Invent. math.
DOI 10.1007/s00222-017-0749-x

Modularity lifting beyond the Taylor–Wiles method

Frank Calegari[1] · **David Geraghty**[2]

Received: 16 July 2012 / Accepted: 29 June 2017

Abstract We prove new modularity lifting theorems for p-adic Galois representations in situations where the methods of Wiles and Taylor–Wiles do not apply. Previous generalizations of these methods have been restricted to situations where the automorphic forms in question contribute to a single degree of cohomology. In practice, this imposes several restrictions—one must be in a Shimura variety setting and the automorphic forms must be of regular weight at infinity. In this paper, we essentially show how to remove these restrictions. Our most general result is a modularity lifting theorem which, on the automorphic side, applies to automorphic forms on the group $\mathrm{GL}(n)$ over a general number field; it is contingent on a conjecture which, in particular, predicts the existence of Galois representations associated to torsion classes in the cohomology of the associated locally symmetric space. We show that if this conjecture holds, then our main theorem implies the following: if E is an elliptic curve over an arbitrary number field, then E is potentially automorphic and

The first author was supported in part by NSF Career Grant DMS-0846285, NSF Grant DMS-1404620, NSF Grant DMS-1648702, NSF Grant DMS-1701703, and a Sloan Foundation Fellowship; the second author was supported in part by NSF Grant DMS-1440703; both authors were supported in part by NSF Grant DMS-0841491.

✉ Frank Calegari
fcale@math.uchicago.edu

David Geraghty
djgeraghty@gmail.com

[1] The University of Chicago, 5734 S University Ave, Chicago, IL 60637, USA

[2] Department of Mathematics, Carney Hall, Boston College, Chestnut Hill, MA 02467-3806, USA

Published online: 27 July 2017

satisfies the Sato–Tate conjecture. In addition, we also prove some unconditional results. For example, in the setting of $\mathrm{GL}(2)$ over $\mathbf{Q}$, we identify certain minimal global deformation rings with the Hecke algebras acting on spaces of p-adic Katz modular forms of weight 1. Such algebras may well contain p-torsion. Moreover, we also completely solve the problem (for p odd) of determining the multiplicity of an irreducible modular representation $\overline{\rho}$ in the Jacobian $J_1(N)$, where N is the minimal level such that $\overline{\rho}$ arises in weight two.

Mathematics Subject Classification 11F33 · 11F80

Contents

1 Introduction

In this paper, we prove a new kind of modularity lifting theorem for p-adic Galois representations. Previous generalizations of the work of Wiles [1] and Taylor–Wiles [2] have (essentially) been restricted to circumstances where the automorphic forms in question arise from the middle degree cohomology of Shimura varieties. In particular, such approaches ultimately rely on a "numerical coincidence" (see the introduction to [3]) which does not hold in general, and does not hold in particular for $\mathrm{GL}(2)/F$ if F is not totally real. A second requirement of these generalizations is that the Galois representations in question are *regular* at ∞, that is, have distinct Hodge–Tate weights for all $v|p$. Our approach, in contrast, does not *a priori* require either such assumption.

When considering questions of modularity in more general contexts, there are two issues that need to be overcome. The first is that there do not seem to be "enough" automorphic forms to account for all the Galois representations. In [4–6], the suggestion is made that one should instead consider *integral* cohomology, and that the torsion occurring in these cohomology groups may account for the missing automorphic forms. In order to make this approach work, one needs to show that there is "enough" torsion. This is the problem that we solve in some cases. A second problem is the lack of Galois representations attached to these integral cohomology classes. In particular, our methods require Galois representations associated to torsion classes which do not necessarily lift to characteristic zero, where one might hope to apply the recent results of [7]. We do not resolve the problems of constructing Galois representations in this paper, and instead, our results are contingent on a conjecture which predicts that there exists a map from a suitable deformation ring $R^{\min}$ to a Hecke algebra $\mathbf{T}$. In a recent preprint, Scholze [8] has constructed Galois representations associated to certain torsion classes. If one can show that these Galois representations satisfy certain local-global compatibility conditions (including showing that the Galois representations associated to cohomology classes on which U_v for $v|p$ is invertible are reducible after restriction to the decomposition group at v), then our modularity lifting theorems for imaginary quadratic fields would be unconditional. There *are* contexts, however, in which the existence of Galois representations is known; in these cases we can produce unconditional results. In principle, our method currently applies in two contexts:

Betti To Galois representations conjecturally arising from tempered π of cohomological type associated to G, where G is reductive with a maximal compact K, maximal $\mathbf{Q}$-split torus A, and $l_0 = \mathrm{rank}(G) - \mathrm{rank}(K) - \mathrm{rank}(A)$ is arbitrary,

Coherent To Galois representations conjecturally arising from tempered π associated to G, where (G, X) is a Shimura variety over a totally real field F, and such that π_v is a non-degenerate limit of discrete series at ℓ_0 infinite places and a discrete series at all other infinite places.

In practice, however, what we really need is that (after localizing at a suitable maximal ideal $\mathfrak{m}$ of the Hecke algebra $\mathbf{T}$) the cohomology is concentrated in $l_0 + 1$ consecutive degrees. (This is certainly true of the tempered representations which occur in Betti cohomology. According to [9], the range of cohomological degrees to which they occur has length $l_0 + 1$. In the coherent case, the value of l_0 will depend on the infinity components π_v allowed. That tempered representations occur in a range of length l_0, then follows from [10, Theorems 3.4 and 3.5] together with knowledge of L-packets at infinite primes). The specialization of our approach to the case $\ell_0 = 0$ exactly recovers the usual Taylor–Wiles method.

The following results are a sample of what can be shown by these methods in case **Betti**, assuming (Conjecture B of Sect. 9.3) the existence of Galois representations in appropriate degrees satisfying the expected properties.

Theorem 1.1 *Assume Conjecture* B. *Let F be any number field, and let E be an elliptic curve over F. Then the following hold*:

(1) *E is potentially modular.*
(2) *The Sato–Tate conjecture is true for E.*

The proof of Theorem 1.1 relies on the following ingredients. The first ingredient consists of the usual techniques in modularity lifting (the Taylor–Wiles–Kisin method) as augmented by Taylor's Ihara's Lemma avoidance trick [11]. The second ingredient is to observe that these arguments *continue to hold* in a more general situation, *provided* that one can show that there is "enough" cohomology. Ultimately, this amounts to giving a lower bound on the depth of certain patched Hecke modules. Finally, one can obtain such a lower bound by a commutative algebra argument, *assuming* that the relevant cohomology occurs only in a certain range of length l_0. Conjecture B amounts to assuming both the existence of Galois representations together with the vanishing of cohomology (localized at an appropriate $\mathfrak{m}$) outside a given range. We deduce Theorem 1.1 from a more general modularity lifting theorem, see Theorem 5.16.

The following result is a sample of what can be shown by these methods in case **Betti** assuming only Conjecture A concerning the existence of Galois representations for arithmetic lattices in $\mathrm{GL}_2(\mathcal{O}_F)$ for an imaginary quadratic field F. Unlike Conjecture B, it appears that Conjecture A may well be quite tractable in light of the work of [8]. Let $\mathcal{O}$ denote the ring of integers in a finite extension of $\mathbf{Q}_p$, let ϖ be a uniformizer of $\mathcal{O}$, and let $\mathcal{O}/\varpi = k$ be the residue field. Say that a representation $\rho : G_F \rightarrow \mathrm{GL}_2(\mathcal{O})$ is semi-stable if $\rho|I_v$ is unipotent for all finite v not dividing p, and semi-stable in the sense of Fontaine [12] if $v|p$. Furthermore, for $v|p$, we say that $\rho|D_v$ is finite flat if for all $n \geq 1$ each finite quotient $\rho|D_v \mod \varpi^n$ is the generic fiber of a finite flat $\mathcal{O}$-group scheme, and ordinary if $\rho|D_v$ is conjugate to a representation of the form

$$\begin{pmatrix} \epsilon\chi_1 & * \\ 0 & \chi_2 \end{pmatrix}$$

where χ_1 and χ_2 are unramified and ϵ is the cyclotomic character.

Theorem 1.2 *Assume Conjecture A. Let $F/\mathbf{Q}$ be an imaginary quadratic field. Let $p \geq 3$ be unramified in F. Let*

$$\rho : G_F \rightarrow \mathrm{GL}_2(\mathcal{O})$$

be a continuous semi-stable Galois representation with cyclotomic determinant unramified outside finitely many primes. Let $\overline{\rho} : G_F \rightarrow \mathrm{GL}_2(k)$ denote the mod-ϖ reduction of ρ. Suppose that

(1) *If $v|p$, the representation $\rho|D_v$ is either finite flat or ordinary.*
(2) *The restriction of $\overline{\rho}$ to $G_{F(\zeta_p)}$ is absolutely irreducible.*
(3) *$\overline{\rho}$ is modular of level $N(\overline{\rho})$, where $N(\overline{\rho})$ is the product of the usual prime-to-p Artin conductor and the primes $v|p$ where $\overline{\rho}$ is not finite-flat.*
(4) *ρ is minimally ramified.*

Then ρ is modular, that is, there exists a regular algebraic cusp form π for $\mathrm{GL}(2)/F$ such that $L(\rho, s) = L(\pi, s)$.

It is important to note that the condition (3) is only a statement about the existence of a mod-p cohomology class of level $N(\overline{\rho})$, not the existence of a characteristic zero lift. This condition is the natural generalization of Serre's conjecture.

It turns out that—even assuming Conjecture A—this is not enough to prove that all minimal semi-stable elliptic curves over F are modular. Even though the Artin conjecture for finite two-dimensional solvable representations of G_F is known, there are no obvious congruences between eigenforms arising from Artin L-functions and cohomology classes over F. (Over $\mathbf{Q}$, this arose from the happy accident that classical weight one forms could be interpreted via coherent cohomology). One class of mod-p Galois representations known to satisfy (3) are the restrictions of odd Galois representations $\overline{\rho} : G_{\mathbf{Q}} \rightarrow \mathrm{GL}_2(k)$ to G_F. One might imagine that the minimality condition is a result of the lack of Ihara's lemma; however, Ihara's lemma and level raising are known for $\mathrm{GL}(2)/F$ (see [6]). The issue arises because there is no analogue of Wiles' numerical criterion for Gorenstein rings of dimension zero.

We deduce Theorem 1.2 from the following more general result.

Theorem 1.3 *Assume conjecture A. Let $F/\mathbf{Q}$ be an imaginary quadratic field. Let $p \geq 3$ be unramified in F. Let*

$$\overline{\rho} : G_F \rightarrow \mathrm{GL}_2(k)$$

be a continuous representation with cyclotomic determinant, and suppose that:

(1) *If $v|p$, the representation $\overline{\rho}|D_v$ is either finite flat or ordinary.*
(2) *$\overline{\rho}$ is modular of level $N = N(\overline{\rho})$.*
(3) *$\overline{\rho}|G_{F(\zeta_p)}$ is absolutely irreducible.*
(4) *If $\overline{\rho}$ is ramified at x where $N_{F/\mathbf{Q}}(x) \equiv -1 \mod p$, then either $\overline{\rho}|D_x$ is reducible or $\overline{\rho}|I_x$ is absolutely irreducible.*

Let $R^{\min}$ denote the minimal finite flat (respectively, ordinary) deformation ring of $\overline{\rho}$ with cyclotomic determinant. Let $\mathbf{T}_{\mathfrak{m}}$ be the algebra of Hecke operators acting on $H_1(Y_0(N), \mathcal{O})$ localized at the maximal ideal corresponding to $\overline{\rho}$. Then there is an isomorphism:

$$R^{\min} \xrightarrow{\sim} \mathbf{T}_{\mathfrak{m}},$$

and there exists an integer $\mu \geq 1$ such that $H_1(Y_0(N), \mathcal{O})_{\mathfrak{m}}$ is free of rank μ as a $\mathbf{T}_{\mathfrak{m}}$-module. If $H_1(Y_0(N), \mathcal{O})_{\mathfrak{m}} \otimes \mathbf{Q} \neq 0$, then $\mu = 1$. If $\dim(\mathbf{T}_{\mathfrak{m}}) = 0$, then $\mathbf{T}_{\mathfrak{m}}$ is a complete intersection.

Note that condition 4—the non-existence of "vexing primes" x such that $N_{F/\mathbf{Q}}(x) \equiv -1 \mod p$—is already a condition that arises in the original paper of Wiles [1]. It could presumably be removed by making the appropriate modifications as in either [13,14] or [15] and making the corresponding modifications to Conjecture A.

Our results are obtained by applying a modification of Taylor–Wiles to the Betti cohomology of arithmetic manifolds. In such a context, it seems difficult to construct Galois representations whenever $l_0 \neq 0$. Following [16,17], however, we may also apply our methods to the *coherent* cohomology of Shimura varieties, where Galois representations are more readily available. In contexts where the underlying automorphic forms π are discrete series at infinity, one expects (and in many cases can prove, see [18]) that the integral coherent cohomology localized at a suitably generic maximal ideal $\mathfrak{m}$ of $\mathbf{T}$ vanishes outside the middle degree. If π_∞ is a *limit* of discrete series, however, (so that we are in case **Coherent**) then the cohomology of the associated coherent sheaf can sometimes be shown to be non-zero in exactly in the expected number of degrees, in which case our methods apply. In particular, *a priori*, the **Coherent** case appears more tractable, since there are available methods for constructing Galois representations to coherent cohomology classes in low degree [19,20]. However, our methods *still* lead to open conjectures concerning the existence of Galois representations, since the usual methods for constructing representations on torsion classes (using congruences) only work with Hecke actions on $H^0(X, \mathscr{E})$ rather than $H^i(X, \mathscr{E})$ for $i > 0$, and we require Galois representations coming from the latter groups.

In this paper, we confine our discussion of the general **Coherent** case to addressing the problem of constructing suitable complexes (see Sect. 7.2). We expect, however, that our methods may be successfully applied to prove unconditional modularity lifting theorems in a number of interesting cases in small rank. The most well known example of such a situation is the case of classical modular forms of weight 1. Such modular forms contribute to the cohomology of $H^0(X_1(N), \omega)$ *and* $H^1(X_1(N), \omega)$ in characteristic zero, where $X_1(N)$ is a modular curve, and ω is the usual pushforward $\pi_*\omega_{\mathcal{E}/X_1(N)}$ of the relative dualizing sheaf along the universal generalized elliptic curve. Working over $\mathbf{Z}_p$ for some prime $p \nmid N$, the group $H^0(X_1(N), \omega)$ is torsion free, but $H^1(X_1(N), \omega)$ is not torsion free in general, as predicted by Serre and confirmed by Mestre (Appendix A of [21]). In order to deal with vexing primes, we introduce a vector bundle $\mathcal{L}_\sigma$ which plays the role of the locally constant sheaf $\mathcal{F}_M$ of section Sect. 6 of [15]—see Sect. 3.9 for details. We also introduce a curve X_U which sits in the sequence $X_1(M) \to X_U \to X_0(M)$ for some M dividing the Serre conductor of $\overline{\rho}$ and such that the first map has p-power degree. Note that if $\overline{\rho}$ has no vexing primes, then $\mathcal{L}_\sigma$ is trivial of rank 1, and if $\overline{\rho}$ is not ramified at any primes congruent to $1 \mod p$, then $X_U = X_1(N)$. In this context, we prove the following result:

Theorem 1.4 *Suppose that $p \geq 3$. Let $\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(k)$ be an odd continuous irreducible Galois representation of Serre level N. Assume that $\overline{\rho}$ is unramified at p. Let $R^{\min}$ denote the universal minimal unramified-at-p deformation ring of $\overline{\rho}$. Then there exists a quotient X_U of $X_1(N)$ and a vector bundle $\mathcal{L}_\sigma$ on X_U such that if $\mathbf{T}$ denotes the Hecke algebra of $H^1(X_U, \omega\otimes\mathcal{L}_\sigma)$, there is an isomorphism*

$$R^{\min} \xrightarrow{\sim} \mathbf{T}_{\mathfrak{m}}$$

where $\mathfrak{m}$ is the maximal ideal of $\mathbf{T}$ corresponding to $\overline{\rho}$. Moreover, $H^1(X_U, \omega\otimes \mathcal{L}_\sigma)_{\mathfrak{m}}$ is free as a $\mathbf{T}_{\mathfrak{m}}$-module.

Note that even the fact that there *exists* a surjective map from $R^{\min}$ to $\mathbf{T}_{\mathfrak{m}}$ is non-trivial, and requires us to prove a local–global compatibility result for Galois representations associated to Katz modular forms of weight one over any $\mathbf{Z}_p$-algebra (see Theorem 3.11). We immediately deduce from Theorem 1.4 the following:

Corollary 1.5 *Suppose that* $p \geq 3$. *Suppose also that* $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathcal{O})$ *is a continuous representation satisfying the following conditions.*

(1) *For all primes* v, *either* $\rho(I_v) \xrightarrow{\sim} \overline{\rho}(I_v)$ *or* $\dim(\rho^{I_v}) = \dim(\overline{\rho}^{I_v}) = 1$.
(2) $\overline{\rho}$ *is odd and irreducible.*
(3) ρ *is unramified at* p.

Then ρ *is modular of weight one.*

It is instructive to compare this theorem and the corollary to the main theorem of Buzzard–Taylor [22] (see also [23]). Note that the hypothesis in that paper that $\overline{\rho}$ is modular is no longer necessary, following the proof of Serre's conjecture [24]. In both cases, if ρ is a deformation of $\overline{\rho}$ to a field of characteristic zero, we deduce that ρ is modular of weight one, and hence has finite image. The method of [22] applies in a non-minimal situation, but it requires the hypothesis that $\overline{\rho}(\mathrm{Frob}_p)$ has distinct eigenvalues. Moreover, it has the disadvantage that it only gives an identification of reduced points on the generic fibre (equivalently, that $R^{\min}[1/p]^{\mathrm{red}} = \mathbf{T}_{\mathfrak{m}}[1/p]$, although from this by class field theory—see Lemma 4.14 and the subsequent remarks after the proof—one may deduce that $R^{\min}[1/p] = \mathbf{T}_{\mathfrak{m}}[1/p]$), and says nothing about the torsion structure of $H^1(X_1(N), \omega)$. Contrastingly, we may deduce the following result:

Corollary 1.6 *Suppose that* $p \geq 3$. *Let* $\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(k)$ *be odd, continuous, irreducible, and unramified at* p. *Let* $(A, \mathfrak{m})$ *denote a complete local Noetherian* $\mathcal{O}$*-algebra with residue field* k *and* $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(A)$ *a minimal deformation of* $\overline{\rho}$. *Then* ρ *has finite image.*

This gives the first results towards Boston's strengthening of the Fontaine–Mazur conjecture for representations unramified at p (See [25], Conjecture 2).

It is natural to ask whether our results can be modified using Kisin's method to yield modularity lifting results in non-minimal level. Although the formalism of this method can be adapted to our context, there is a genuine difficultly in proving that the support of $\mathrm{Spec}(\mathbf{T}_\infty[1/p])$ hits each of the components of $\mathrm{Spec}(R_\infty[1/p])$ whenever the latter has more than one component. In certain situations, we may apply Taylor's trick [11], but this can not be made to work in general. However, suppose one replaces the "minimal" condition away from p with the following condition:

- If ρ is special at $x \nmid p$, and $\overline{\rho}$ is unramified at x, then $x \equiv 1 \mod p$.

In this context our methods should yield that the deformation ring R acts nearly faithfully on $H^1(X_1(M), \omega)_{\mathfrak{m}}$ for an appropriate M. This is sufficient for applications to the conjectures of Fontaine–Mazur and Boston.

In the process of proving our main result, we also completely solve the problem (for p odd) of determining the multiplicity of an irreducible modular representation $\overline{\rho}$ in the Jacobian $J_1(N^*)[\mathfrak{m}]$, where N^* is the minimal level such that $\overline{\rho}$ arises in weight two. In particular, we prove that when $\overline{\rho}$ is unramified at p and $\overline{\rho}(\mathrm{Frob}_p)$ is a scalar, then the multiplicity of $\overline{\rho}$ is two (see Theorem 4.8). (In all other cases, the multiplicity was already known to be one—and in the exceptional cases we consider, the multiplicity was also known to be ≥ 2).

Finally, we outline here the structure of the paper, which has two parts. In Part 1, we treat the case where $l_0 = 1$ in two specific instances—namely, the case of classical modular forms of weight 1 (Sect. 3) and the case of automorphic forms on $\mathrm{GL}(2)$ over a quadratic imaginary field that contribute to the Betti cohomology (Sect. 5). The ideas from commutative algebra and the abstract Taylor–Wiles patching method necessary to treat these two situations are developed in Sect. 2. The 'multiplicity two' result mentioned above is proved in Sect. 4.

In Part 2, we treat the case of general l_0. In contrast to Part 1, we only treat the Betti case in detail (more specifically, we consider the Betti cohomology of the locally symmetric spaces associated to $\mathrm{GL}(n)$ over a general number field). Section 6 contains the results from commutative algebra and the abstract Taylor–Wiles style patching result that underlie our approach to the case of general l_0. These techniques are more 'derived' in nature than the techniques that treat $l_0 = 0$ or 1, and in particular rely on the existence of complexes which compute cohomology and satisfy various desirable properties. The existence of such complexes is proved in Sect. 7 (in both contexts—Betti cohomology and coherent cohomology). In Sect. 8, we consider the Galois deformation side of our arguments. In Sect. 9, we consider cohomology and Hecke algebras. This section contains Conjecture B on the existence of Galois representations as well our main modularity lifting theorem. Finally, Sect. 10 contains the proof of Theorem 1.1 above.

Notation

In this paper, we fix a prime $p \geq 3$ and let $\mathcal{O}$ denote the ring of integers in a finite extension K of $\mathbf{Q}_p$. We let ϖ denote a uniformizer in $\mathcal{O}$ and let $k = \mathcal{O}/\varpi$ be the residue field. We denote by $\mathcal{C}_{\mathcal{O}}$ the category of complete Noetherian local $\mathcal{O}$-algebras with residue field k. The homomorphisms in $\mathcal{C}_{\mathcal{O}}$ are the continuous $\mathcal{O}$-algebra homomorphisms. If G is a group and $\chi : G \to k^\times$ is a character, we denote by $\langle \chi \rangle : G \to \mathcal{O}^\times$ the Teichmüller lift of χ.

If F is a field, we let G_F denote the Galois group $\mathrm{Gal}(\overline{F}/F)$ for some choice of algebraic closure $\overline{F}/F$. We let $\epsilon : G_F \to \mathbf{Z}_p^\times$ denote the p-adic cyclotomic character. If F is a number field and v is a prime of F, we let $\mathcal{O}_v$ denote the ring of integers in the completion of F at v and we let π_v denote a uniformizer in $\mathcal{O}_v$. We denote G_{F_v} by G_v and let $I_v \subset G_v$ be the inertia group. We also let $\mathrm{Frob}_v \in G_v/I_v$ denote the *arithmetic* Frobenius. We let $\mathrm{Art} : F_v^\times \to W_{F_v}^{\mathrm{ab}}$ denote the local Artin map, normalized to send uniformizers to geometric Frobenius lifts. We will also sometimes denote the decomposition group at v by D_v. If R is a topological ring and $\alpha \in R^\times$, we let $\lambda(\alpha) : G_v \to R^\times$ denoted the continuous unramified character which sends Frob_v to α, when such a character exists. We let $\mathbf{A}_F$ and $\mathbf{A}_F^\infty$ denote the adeles and finite adeles of F respectively. If $F = \mathbf{Q}$, we simply write $\mathbf{A}$ and $\mathbf{A}^\infty$.

If P is a bounded complex of S-modules for some ring S, then we let $H^*(P) = \oplus_i H^i(P)$. Any map $H^*(P) \to H^*(P)$ will be assumed to be degree preserving. If R is a ring, by a *perfect complex of* R*-modules* we mean a bounded complex of finitely generated projective R-modules.

If R is a local ring, we will sometimes denote the maximal ideal of R by $\mathfrak{m}_R$.

Part 1. l_0 **equals** 1

2 Some commutative algebra I

This section contains one of the main new technical innovations of this paper. The issue, as mentioned in the introduction, is to show that there are *enough* modular Galois representations. This involves showing that certain modules H_N (consisting of modular forms) for the group rings $S_N := \mathcal{O}[(\mathbf{Z}/p^N\mathbf{Z})^q]$ compile, in a Taylor–Wiles patching process, to form a module of codimension one over the completed group ring $S_\infty := \mathcal{O}[(\mathbf{Z}_p)^q]$. The problem then becomes to find a suitable notion of "codimension one" for modules over a local ring that

(1) is well behaved for non-reduced quotients of power series rings over $\mathcal{O}$ (like S_N),
(2) can be established for the spaces H_N in question,
(3) compiles well in a Taylor–Wiles system.

It turns out that the correct notion is that of being "balanced", a notion defined below. When $l_0 > 1$, we shall ultimately be required to patch more information than simply the modules H_N; rather, we shall patch entire complexes (see Sect. 6).

2.1 Balanced modules

Let S be a Noetherian local ring with residue field k and let M be a finitely generated S-module.

Definition 2.1 We define the *defect* $d_S(M)$ of M to be

$$\begin{aligned} d_S(M) &= \dim_k \mathrm{Tor}^0_S(M,k) - \dim_k \mathrm{Tor}^1_S(M,k) \\ &= \dim_k M/\mathfrak{m}_S M - \dim_k \mathrm{Tor}^1_S(M,k). \end{aligned}$$

Let

$$\cdots \to P_i \to \cdots \to P_1 \to P_0 \to M \to 0$$

be a (possibly infinite) resolution of M by finite free S-modules. Assume that the image of P_i in P_{i-1} is contained in $\mathfrak{m}_S P_{i-1}$ for each $i \geq 1$. (Such resolutions always exist and are often called 'minimal'). Let r_i denote the rank of P_i. Tensoring the resolution over S with k we see that $P_i/\mathfrak{m}_S P_i \cong \mathrm{Tor}^i_S(M,k)$ and hence that $r_i = \dim_k \mathrm{Tor}^i_S(M,k)$.

Definition 2.2 We say that M is *balanced* if $d_S(M) \geq 0$.

If M is balanced, then we see that it admits a presentation

$$S^d \to S^d \to M \to 0$$

with $d = \dim_k M/\mathfrak{m}_S M$.

2.2 Patching

We establish in this section an abstract Taylor–Wiles style patching result which may be viewed as an analogue of Theorem 2.1 of [26]. This result will be one of the key ingredients in the proofs of our main theorems.

Proposition 2.3 *Suppose that*

(1) *R is an object of $\mathcal{C}_\mathcal{O}$ and H is a finite R-module which is also finite over $\mathcal{O}$;*
(2) *$q \geq 1$ is an integer, and for each integer $N \geq 1$, $S_N := \mathcal{O}[\Delta_N]$ with $\Delta_N := (\mathbf{Z}/p^N\mathbf{Z})^q$;*
(3) *$R_\infty := \mathcal{O}[x_1, \ldots, x_{q-1}]$;*
(4) *for each $N \geq 1$, $\phi_N : R_\infty \twoheadrightarrow R$ is a surjection in $\mathcal{C}_\mathcal{O}$ and H_N is an $R_\infty \otimes_\mathcal{O} S_N$-module.*
(5) *For each $N \geq 1$ the following conditions are satisfied*

(a) *the image of* S_N *in* $\mathrm{End}_{\mathcal{O}}(H_N)$ *is contained in the image of* R_∞ *and moreover, the image of the augmentation ideal of* S_N *in* $\mathrm{End}_{\mathcal{O}}(H_N)$ *is contained in the image of* $\ker(\phi_N)$*;*
(b) *there is an isomorphism* $\psi_N : (H_N)_{\Delta_N} \xrightarrow{\sim} H$ *of* R_∞*-modules (where* R_∞ *acts on* H *via* ϕ_N*);*
(c) H_N *is finite and balanced over* S_N *(see Definition 2.2).*

Then H *is a free* R*-module.*

Proof Let $S_\infty = \mathcal{O}[(\mathbf{Z}_p)^q]$ and let $\mathfrak{a}$ denote the augmentation ideal of S_∞ (that is, the kernel of the homomorphism $S_\infty \twoheadrightarrow \mathcal{O}$ which sends each element of $(\mathbf{Z}_p)^q$ to 1). For each $N \geq 1$, let $\mathfrak{a}_N$ denote the kernel of the natural surjection $S_\infty \twoheadrightarrow S_N$ and let $\mathfrak{b}_N$ denote the open ideal of S_∞ generated by ϖ^N and $\mathfrak{a}_N$. Let $d = \dim_k(H/\varpi H)$. We may assume that $d > 0$ since otherwise $H = \{0\}$ and the result is trivially true. Choose a sequence of open ideals $(\mathfrak{d}_N)_{N \geq 1}$ of R such that

- $\mathfrak{d}_N \supset \mathfrak{d}_{N+1}$ for all $N \geq 1$;
- $\cap_{N \geq 1} \mathfrak{d}_N = (0)$;
- $\varpi^N R \subset \mathfrak{d}_N \subset \varpi^N R + \mathrm{Ann}_R(H)$ for all N.

(For example, one can take $\mathfrak{d}_N$ to be the ideal generated by ϖ^N and $\mathrm{Ann}_R(H)^N$. These are open ideals since $R/\mathrm{Ann}_R(H) \subset \mathrm{End}_{\mathcal{O}}(H)$ is finite as an $\mathcal{O}$-module).

Define a *patching datum of level* N to be a 4-tuple (ϕ, X, ψ, P) where

- $\phi : R_\infty \twoheadrightarrow R/\mathfrak{d}_N$ is a surjection in $\mathcal{C}_{\mathcal{O}}$;
- X is an $R_\infty \widehat{\otimes}_{\mathcal{O}} S_\infty$-module such that the action of S_∞ on X factors through $S_\infty/\mathfrak{b}_N$ and X is finite over S_∞;
- $\psi : X/\mathfrak{a}X \xrightarrow{\sim} H/\varpi^N H$ is an isomorphism of R_∞ modules (where R_∞ acts on $H/\varpi^N H$ via ϕ);
- P is a presentation

$$(S_\infty/\mathfrak{b}_N)^d \to (S_\infty/\mathfrak{b}_N)^d \to X \to 0.$$

We say that two such 4-tuples (ϕ, X, ψ, P) and (ϕ', X', ψ', P') are isomorphic if

- $\phi = \phi'$;
- there is an isomorphism $X \xrightarrow{\sim} X'$ of $R_\infty \widehat{\otimes}_{\mathcal{O}} S_\infty$ modules compatible with ψ and ψ', and with the presentations P and P'.

We note that there are only finitely many isomorphism classes of patching data of level N. (This follows from the fact that R_∞ and S_∞ are topologically finitely generated). If D is a patching datum of level N and $1 \leq N' \leq N$, then

D gives rise to patching datum of level N' in an obvious fashion. We denote this datum by D mod N'.

For each pair of integers (M, N) with $M \geq N \geq 1$, we define a patching datum $D_{M,N}$ of level N as follows: the statement of the proposition gives a homomorphism $\phi_M : R_\infty \twoheadrightarrow R$ and an $R_\infty \otimes_{\mathcal{O}} S_M$-module H_M. We take

- ϕ to be the composition $R_\infty \twoheadrightarrow R \twoheadrightarrow R/\mathfrak{d}_N$;
- X to be $H_M/\mathfrak{b}_N$;
- $\psi : X/\mathfrak{a}X \xrightarrow{\sim} H/\mathfrak{b}_N$ to be the reduction modulo ϖ^N of the given isomorphism $\psi_M : H_M/\mathfrak{a}H_M \xrightarrow{\sim} H$;
- P to be any choice of presentation

$$(S_\infty/\mathfrak{b}_N)^d \to (S_\infty/\mathfrak{b}_N)^d \to X \to 0.$$

(The facts that $H_M/\mathfrak{a}H_M \xrightarrow{\sim} H$ and $d_{S_M}(H_M) \geq 0$ imply that such a presentation exists).

Since there are finitely many patching data of each level $N \geq 1$, up to isomorphism, we can find a sequence of pairs $(M_i, N_i)_{i\geq 1}$ such that

- $M_i \geq N_i$, $M_{i+1} > M_i$, and $N_{i+1} > N_i$ for all i;
- $D_{M_{i+1},N_{i+1}}$ mod N_i is isomorphic to D_{M_i,N_i} for all $i \geq 1$.

For each $i \geq 1$, we write $D_{M_i,N_i} = (\phi_i, X_i, \psi_i, P_i)$ and we fix an isomorphism between the modules $X_{i+1}/\mathfrak{b}_{N_i} X_{i+1}$ and X_i giving rise to an isomorphism between $D_{M_{i+1},N_{i+1}}$ mod N_i and D_{M_i,N_i}. We define

- $\phi_\infty : R_\infty \twoheadrightarrow R$ to be the inverse limit of the ϕ_i;
- $X_\infty := \varprojlim_i X_i$ where the map $X_{i+1} \to X_i$ is the composition $X_{i+1} \twoheadrightarrow X_{i+1}/\mathfrak{b}_{N_i} \xrightarrow{\sim} X_i$;
- ψ_∞ to be the isomorphism of R_∞-modules $X_\infty/\mathfrak{a}X_\infty \xrightarrow{\sim} H$ (where R_∞ acts on H via ϕ_∞) arising from the isomorphisms ψ_i;
- P_∞ to be the presentation

$$S_\infty^d \to S_\infty^d \to X_\infty \to 0$$

obtained from the P_i. (Exactness follows from the Mittag–Leffler condition).

Then X_∞ is an $R_\infty \widehat{\otimes}_{\mathcal{O}} S_\infty$-module, and the image of S_∞ in $\mathrm{End}_{\mathcal{O}}(X_\infty)$ is contained in the image of R_∞. (By condition 5a, the image of S_∞ in each $\mathrm{End}_{\mathcal{O}}(X_i)$ is contained in the image of R_∞. The same containment of images then holds in each $\mathrm{Hom}_{\mathcal{O}}(X_\infty, X_i)$ and hence in $\mathrm{End}_{\mathcal{O}}(X_\infty) = \varprojlim_i \mathrm{Hom}_{\mathcal{O}}(X_\infty, X_i)$). It follows that X_∞ is a finite R_∞-module. Since S_∞ is formally smooth over $\mathcal{O}$, we can and do choose a homomorphism $\iota : S_\infty \to R_\infty$ in $\mathcal{C}_{\mathcal{O}}$, compatible with the actions of S_∞ and R_∞ on X_∞.

Since $\dim_{S_\infty}(X_\infty) = \dim_{R_\infty}(X_\infty)$ and $\dim R_\infty < \dim S_\infty$, we deduce that $\dim_{S_\infty}(X_\infty) < \dim S_\infty$. It follows that the first map $S_\infty^d \to S_\infty^d$ in the presentation P_∞ is injective. (Denote the kernel by K. If $K \neq (0)$, then $K \otimes_{S_\infty} \mathrm{Frac}(S_\infty) \neq (0)$ and hence $X_\infty \otimes_{S_\infty} \mathrm{Frac}(S_\infty) \neq (0)$, which is impossible). We see that P_∞ is a minimal projective resolution of X_∞, and by the Auslander–Buchsbaum formula, we deduce that $\mathrm{depth}_{S_\infty}(X_\infty) = \dim(S_\infty) - 1$. Since $\mathrm{depth}_{R_\infty}(X_\infty) = \mathrm{depth}_{S_\infty}(X_\infty)$, it follows that $\mathrm{depth}_{R_\infty}(X_\infty) = \dim(R_\infty)$, and applying the Auslander–Buchsbaum formula again, we deduce that X_∞ is free over R_∞. Using this and the second part of condition (5a), we also deduce that $\iota(\mathfrak{a}) \subset \ker(\phi_\infty)$.

Finally, the existence of the isomorphism $\psi_\infty : X_\infty/\mathfrak{a}X_\infty \xrightarrow{\sim} H$ tells us that H is free over $R_\infty/\iota(\mathfrak{a})R_\infty$. However, since the action of R_∞ on H also factors through the quotient $R_\infty/\ker(\phi_\infty) = R$ and since $\iota(a) \subset \ker(\phi_\infty)$, we deduce that $R_\infty/\iota(\mathfrak{a})R_\infty \cong R$ and that R acts freely on H. □

3 Weight one forms

3.1 Deformations of Galois representations

Let

$$\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(k)$$

be a continuous, odd, absolutely irreducible Galois representation. Let us suppose that $\overline{\rho}|G_p$ is unramified; this implies that $\overline{\rho}$ remains absolutely irreducible when restricted to $G_{\mathbf{Q}(\zeta_p)}$. Let $S(\overline{\rho})$ denote the set of primes of $\mathbf{Q}$ at which $\overline{\rho}$ is ramified and let $T(\overline{\rho}) \subset S(\overline{\rho})$ be the subset consisting of those primes x such that $x \equiv -1 \mod p$, $\overline{\rho}|G_x$ is irreducible and $\overline{\rho}|I_x$ is reducible. Following Diamond, we call the primes in $T(\overline{\rho})$ *vexing*. We further assume that if $x \in S(\overline{\rho})$ and $\overline{\rho}|G_x$ is reducible, then $\overline{\rho}^{I_x} \neq (0)$. Note that this last condition is always satisfied by a twist of $\overline{\rho}$ by a character unramified outside of $S(\overline{\rho})$.

Let Q denote a finite set of primes of $\mathbf{Q}$ disjoint from $S(\overline{\rho}) \cup \{p\}$. (By abuse of notation, we sometimes use Q to denote the product of primes in Q). For objects R in $\mathcal{C}_{\mathcal{O}}$, a *deformation* of $\overline{\rho}$ to R is a $\ker(\mathrm{GL}_2(R) \to \mathrm{GL}_2(k))$-conjugacy class of continuous lifts $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(R)$ of $\overline{\rho}$. We will often refer to the deformation containing a lift ρ simply by ρ.

Definition 3.1 We say that a deformation $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(R)$ of $\overline{\rho}$ is *minimal outside Q* if it satisfies the following properties:

(1) The determinant $\det(\rho)$ is equal to the Teichmüller lift of $\det(\overline{\rho})$.
(2) If $x \notin Q \cup S(\overline{\rho})$ is a prime of $\mathbf{Q}$, then $\rho|G_x$ is unramified.
(3) If $x \in T(\overline{\rho})$, then $\rho(I_x) \xrightarrow{\sim} \overline{\rho}(I_x)$.

(4) If $x \in S(\overline{\rho}) - T(\overline{\rho})$ and $\overline{\rho}|G_x$ is reducible, then ρ^{I_x} is a rank one direct summand of ρ as an R-module.

If Q is empty, we will refer to such deformations simply as being *minimal*.

Note that condition 2 implies that ρ is unramified at p. The functor that associates to each object R of $\mathcal{C}_{\mathcal{O}}$ the set of deformations of $\overline{\rho}$ to R which are minimal outside Q is represented by a complete Noetherian local $\mathcal{O}$-algebra R_Q. This follows from the proof of Theorem 2.41 of [27]. If $Q = \emptyset$, we will sometimes denote R_Q by $R^{\min}$. Let $H^1_Q(\mathbf{Q}, \mathrm{ad}^0\overline{\rho})$ denote the Selmer group defined as the kernel of the map

$$H^1(\mathbf{Q}, \mathrm{ad}^0\overline{\rho}) \longrightarrow \bigoplus_x H^1(\mathbf{Q}_x, \mathrm{ad}^0\overline{\rho})/L_{Q,x}$$

where x runs over all primes of $\mathbf{Q}$ and

- $L_{Q,x} = H^1(G_x/I_x, (\mathrm{ad}^0\overline{\rho})^{I_x})$ if $x \notin Q$;
- $L_{Q,x} = H^1(\mathbf{Q}_x, \mathrm{ad}^0\overline{\rho})$ if $x \in Q$.

Let $H^1_Q(\mathbf{Q}, \mathrm{ad}^0\overline{\rho}(1))$ denote the corresponding dual Selmer group.

Proposition 3.2 *The reduced tangent space* $\mathrm{Hom}(R_Q/\mathfrak{m}_{\mathcal{O}}, k[\epsilon]/\epsilon^2)$ *of* R_Q *has dimension*

$$\dim_k H^1_Q(\mathbf{Q}, \mathrm{ad}^0\overline{\rho}(1)) - 1 + \sum_{x \in Q} \dim_k H^0(\mathbf{Q}_x, \mathrm{ad}^0\overline{\rho}(1)).$$

Proof The argument is very similar to that of Corollary 2.43 of [27]. The reduced tangent space has dimension $\dim_k H^1_Q(\mathbf{Q}, \mathrm{ad}^0\overline{\rho})$. By Theorem 2.18 of *op. cit.* this is equal to

$$\begin{aligned}&\dim_k H^1_Q(\mathbf{Q}, \mathrm{ad}^0\overline{\rho}(1)) + \dim_k H^0(\mathbf{Q}, \mathrm{ad}^0\overline{\rho}) - \dim_k H^0(\mathbf{Q}, \mathrm{ad}^0\overline{\rho}(1))\\ &\quad + \sum_x (\dim_k L_{Q,x} - \dim_k H^0(\mathbf{Q}_x, \mathrm{ad}^0\overline{\rho})) - 1,\end{aligned}$$

where x runs over all finite places of $\mathbf{Q}$. The final term is the contribution at the infinite place. The second and third terms vanish by the absolute irreducibility of $\overline{\rho}$ and the fact that $\overline{\rho}|G_p$ is unramified. Finally, as in the proof of Corollary 2.43 of *loc. cit.* we see that the contribution at the prime x vanishes if $x \notin Q$, and equals $\dim_k H^0(\mathbf{Q}_x, \mathrm{ad}^0\overline{\rho}(1))$ if $x \in Q$. □

Suppose that $x \equiv 1 \mod p$ and $\overline{\rho}(\mathrm{Frob}_x)$ has distinct eigenvalues for each $x \in Q$. Then $H^0(\mathbf{Q}_x, \mathrm{ad}^0\overline{\rho}(1))$ is one dimensional for $x \in Q$ and the preceding proposition shows that the reduced tangent space of R_Q has dimension

$$\dim_k H^1_Q(\mathbf{Q}, \mathrm{ad}^0\overline{\rho}(1)) - 1 + \#Q.$$

Using this fact and the argument of Theorem 2.49 of [27], we deduce the following result. (We remind the reader that $\overline{\rho}|G_{\mathbf{Q}(\zeta_p)}$ is absolutely irreducible, by assumption).

Proposition 3.3 *Let $q = \dim_k H^1_\emptyset(\mathbf{Q}, \mathrm{ad}^0\overline{\rho}(1))$. Then $q \geq 1$ and for any integer $N \geq 1$ we can find a set Q_N of primes of $\mathbf{Q}$ such that*

(1) $\# Q_N = q$.
(2) $x \equiv 1 \mod p^N$ *for each* $x \in Q_N$.
(3) *For each $x \in Q_N$, $\overline{\rho}$ is unramified at x and $\overline{\rho}(\mathrm{Frob}_x)$ has distinct eigenvalues.*
(4) $H^1_{Q_N}(\mathbf{Q}, \mathrm{ad}^0\overline{\rho}(1)) = (0)$.

In particular, the reduced tangent space of R_{Q_N} has dimension $q-1$ and R_{Q_N} is a quotient of a power series ring over $\mathcal{O}$ in $q-1$ variables.

We note that the calculations on the Galois side are virtually identical to those that occur in Wiles' original paper, with the caveat that the tangent space is of dimension "one less" in our case. On the automorphic side, this -1 will be a reflection of the fact that the Hecke algebras will not (in general) be flat over $\mathcal{O}$ and the modular forms we are interested in will contribute to one extra degree of cohomology.

3.2 Cohomology of modular curves

3.2.1 Modular curves

We begin by recalling some classical facts regarding modular curves. Fix an integer $N \geq 5$ such that $(N, p) = 1$, and fix a squarefree integer Q with $(Q, Np) = 1$. Let $X_1(N)$, $X_1(N; Q)$, and $X_1(NQ)$ denote the modular curves of level $\Gamma_1(N)$, $\Gamma_1(N) \cap \Gamma_0(Q)$, and $\Gamma_1(N) \cap \Gamma_1(Q)$ respectively as smooth proper schemes over $\mathrm{Spec}(\mathcal{O})$. To be precise, we take $X_1(N)$ and $X_1(NQ)$ to be the base change to $\mathrm{Spec}(\mathcal{O})$ of the curves denoted by the same symbols in [28, Proposition 2.1]. Thus, $X_1(NQ)$ represents the functor that assigns to each $\mathcal{O}$-scheme S the set of isomorphism classes of triples (E, α_{NQ}) where E/S is a generalized elliptic curve and $\alpha_{NQ} : \mu_{NQ} \hookrightarrow E[NQ]$ is an embedding of group schemes whose image meets every irreducible component in each geometric fibre. Given such a triple, we can naturally decompose $\alpha_{NQ} = \alpha_N \times \alpha_Q$ into its N and Q-parts. The group $(\mathbf{Z}/NQ\mathbf{Z})^\times$ acts on $X_1(NQ)$ in the following fashion: $a \in (\mathbf{Z}/NQ\mathbf{Z})^\times$ sends a pair (E, α_{NQ}) to $\langle a \rangle(E, \alpha_{NQ}) := (E, a \circ \alpha_{NQ})$. We let $X_1(N; Q)$ be the smooth proper curve over $\mathrm{Spec}(\mathcal{O})$ classifying triples (E, α_N, C_Q) where E is a generalized elliptic curve, $\alpha_N : \mu_N \hookrightarrow E[N]$ is an embedding of group schemes and $C_Q \subset E[Q]$ is a subgroup étale locally isomorphic to $\mathbf{Z}/Q\mathbf{Z}$ and such that the subgroup $Q + \alpha_N(\mu_N)$ of E meets every irreducible component of every geometric fibre of E. Then $X_1(N, Q)$ is the quotient of $X_1(NQ)$ by the action of $(\mathbf{Z}/Q\mathbf{Z})^\times \subset (\mathbf{Z}/NQ\mathbf{Z})^\times$ and the map $X_1(NQ) \to X_1(N; Q)$ is étale; on points it sends (E, α_{NQ}) to $(E, \alpha_N, \alpha_Q(\mu_Q))$.

For any modular curve X over $\mathcal{O}$, let $Y \subset X$ denote the corresponding open modular curve parametrizing genuine elliptic curves. Let $\pi : \mathcal{E} \to X$ denote the universal generalized elliptic curve, and let $\omega := \pi_* \omega_{\mathcal{E}/X}$, where $\omega_{\mathcal{E}/X}$ is the relative dualizing sheaf. Then the Kodaira–Spencer map (see [29], A1.3.17) induces an isomorphism $\omega^{\otimes 2} \simeq \Omega^1_{Y/\mathcal{O}}$ over Y, which extends to an isomorphism $\omega^{\otimes 2} \simeq \Omega^1_{X/\mathcal{O}}(\infty)$, where ∞ is the reduced divisor supported on the cusps. If R is an $\mathcal{O}$-algebra, we let $X_R = X \times_{\mathrm{Spec}\mathcal{O}} \mathrm{Spec} R$. If M is an $\mathcal{O}$-module and $\mathscr{L}$ is a coherent sheaf on X, we let $\mathscr{L}_M$ denote $\mathscr{L} \otimes_{\mathcal{O}} M$.

We now fix a subgroup H of $(\mathbf{Z}/N\mathbf{Z})^\times$. We let X (resp. $X_1(Q)$, resp. $X_0(Q)$) denote[1] the quotient of $X_1(N)$ (resp. $X_1(NQ)$, resp. $X_1(N; Q)$) by the action of H. Note that each of these curves carries an action of $(\mathbf{Z}/N\mathbf{Z})^\times/H$. We assume that H is chosen so that X is the moduli space (rather than the coarse moduli space) of generalized elliptic curves with $\Gamma_H(N) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N) : d \mod N \in H \right\}$-level structure.

3.2.2 Modular forms with coefficients

The map $j : X_{\mathcal{O}/\varpi^m} \to X$ is a closed immersion. If $\mathscr{L}$ is any $\mathcal{O}$-flat sheaf of $\mathcal{O}_X$-modules on X, this allows us to identify $H^0(X_{\mathcal{O}/\varpi^m}, j^*\mathscr{L})$ with $H^0(X, \mathscr{L}_{\mathcal{O}/\varpi^m})$. For such a sheaf $\mathscr{L}$, we may identify $\mathscr{L}_{K/\mathcal{O}}$ with the direct limit $\varinjlim \mathscr{L}_{\mathcal{O}/\varpi^m}$.

3.2.3 Hecke operators

Let $\mathbf{T}^{\mathrm{univ}}$ denote the commutative polynomial algebra over the group ring $\mathcal{O}[(\mathbf{Z}/NQ\mathbf{Z})^\times]$ generated by indeterminates T_x, U_y for $x \nmid pNQ$ prime and $y | Q$ prime. If $a \in (\mathbf{Z}/NQ\mathbf{Z})^\times$, we let $\langle a \rangle$ denote the corresponding element of $\mathbf{T}^{\mathrm{univ}}$. We recall in this section how the Hecke algebra $\mathbf{T}^{\mathrm{univ}}$ acts on coherent cohomology groups.

We have an étale covering map $X_1(Q) \to X_0(Q)$ with Galois group $(\mathbf{Z}/Q\mathbf{Z})^\times$. Let Δ be a quotient of $\Delta_Q := (\mathbf{Z}/Q\mathbf{Z})^\times$ and let $X_\Delta(Q) \to X_0(Q)$ be the corresponding cover. We will define an action of $\mathbf{T}^{\mathrm{univ}}$ on the groups $H^i(X_\Delta(Q), \mathcal{L}_A)$ for A an $\mathcal{O}$-module, $i = 0, 1$ and $\mathcal{L}$ equal to the bundle $\omega^{\otimes n}$

[1] We apologize in advance that this is not entirely consistent with the usual notation for modular curves. The alternative was to adorn the object X with the (fixed throughout) level structure at N coming from $\overline{\rho}$, which the first author felt too notationally cumbersome.

or $\omega^{\otimes n}(-\infty)$. If $\mathcal{C}_\Delta(Q) \subset X_\Delta(Q)$ denotes the divisor of cusps, we will also define an action of $\mathbf{T}^{\text{univ}}$ on $H^0(\mathcal{C}_\Delta(Q), \omega_A^{\otimes n})$.

First of all, if $a \in (\mathbf{Z}/NQ\mathbf{Z})^\times$ and $\mathcal{L}$ denotes either $\omega^{\otimes n}$ or $\omega^{\otimes n}(-\infty)$ on $X_\Delta(Q)$, then we have a natural isomorphism $\langle a\rangle^*\mathcal{L} \xrightarrow{\sim} \mathcal{L}$. We may thus define the operator $\langle a\rangle$ on $H^i(X_\Delta(Q), \mathcal{L}_A)$ as the pull back

$$H^i(X_\Delta(Q), \mathcal{L}_A) \xrightarrow{\langle a\rangle^*} H^i(X_\Delta(Q), \langle a\rangle^*\mathcal{L}_A) = H^i(X_\Delta(Q), \mathcal{L}_A).$$

When $i = 0$, this is just the usual action of the diamond operators (as in [28, Sect. 3], for instance). We define the action of $\langle a\rangle$ on $H^0(\mathcal{C}_\Delta, \omega_A^{\otimes n})$ in the same way, using the fact that $\langle a\rangle$ preserves $\mathcal{C}_\Delta(Q) \subset X_\Delta(Q)$

Now, let x be a prime number which does not divide pNQ. We let $X_\Delta(Q; x)$ denote the modular curve over $\mathcal{O}$ obtained by adding $\Gamma_0(x)$-level structure to $X_\Delta(Q)$ (or equivalently, by taking the quotient of $X_1(NQ; x)$ by the appropriate subgroup of $(\mathbf{Z}/NQ\mathbf{Z})^\times$). We have two finite flat projection maps

$$\pi_i : X_\Delta(Q; x) \to X_\Delta(Q)$$

for $i = 1, 2$. The map π_1 corresponds to the natural forgetful map on open modular curves, extended by 'contraction' to the compactifications. The map π_2 is defined on the open modular curves $Y_\Delta(Q; x) \to Y_\Delta(Q)$ by sending a tuple (E, α_{NQ}, C_x) to the tuple $(E' := E/C_x, \alpha'_{NQ})$ where α'_{NQ} is the level structure on E' obtained from α_{NQ} by composing with the natural isomorphism $E[NQ] \xrightarrow{\sim} E'[NQ]$. The fact that the π_i extend to the compactifications is ensured by [30, Proposition 4.4.3]. We also have the 'Fricke involution' $w_x : X_\Delta(Q; x) \to X_\Delta(Q; x)$ which is defined on the open modular curve $Y_\Delta(X; x)$ by sending a tuple (E, α_{NQ}, C_x) as above to $(E' := E/C_x, \alpha'_{NQ}, E[x]/C_x)$. Note that this is not really an involution, since $w_x^2(E, \alpha_{NQ}, C_x) = (E, x \circ \alpha_{NQ}, C_x) = \langle x\rangle(E, \alpha_{NQ}, C_x)$. We have $\pi_2 = \pi_1 \circ w_x$, and hence $\pi_2^*\omega = (w_x^* \circ \pi_1^*)\omega = w_x^*\omega$. Let $\mathcal{E}$ denote the universal elliptic curve over $Y_\Delta(Q; x)$ and let and let $\mathcal{C}_x \subset \mathcal{E}$ denote the universal subgroup of order x. Let ϕ denote the quotient map

$$\mathcal{E} \longrightarrow \mathcal{E}/\mathcal{C}_x.$$

Then pull back of differentials along ϕ defines a map of sheaves

$$\phi_{12} : \pi_2^*\omega \longrightarrow \pi_1^*\omega$$

over $Y_\Delta(Q; x)$. This map extends over $X_\Delta(Q; x)$ by [30, Proposition 4.4.3].

We define the operator W_x on $H^0(X_\Delta(Q;x), \omega_A^n)$ by setting W_x to be the composite

$$H^0(X_\Delta(Q;x), \omega_A^n) \xrightarrow{w_x^*} H^0(X_\Delta(Q,x), \pi_2^* \omega_A^n) \xrightarrow{\phi_{12}^{\otimes n}} H^i(X_\Delta(Q;x), \omega_A^n).$$

Explicitly, we have:

$$W_x f(E, \alpha_{NQ}, C_x) = \phi^* f(E/C_x, \alpha'_{NQ}, E[x]/C_x).$$

Thus, using the identification $x : E/E[x] \xrightarrow{\sim} E$ together with the fact that we have an equality $w_x^2(E, \alpha_{NQ}, C_x) = \langle x \rangle (E, \alpha_{NQ}, C_x)$, we see that:

$$W_x^2 = x\langle x \rangle.$$

We will use this fact below

We define Hecke operators T_x on $H^i(X_\Delta(Q), \mathcal{L}_A)$, for $\mathcal{L} = \omega^n$, by setting xT_x to be the composition

$$H^i(X_\Delta(Q), \mathcal{L}_A) \to H^i(X_\Delta(Q,x), \pi_2^* \mathcal{L}_A) \stackrel{\phi_{12}^{\otimes n}}{\to} H^i(X_\Delta(Q;x), \pi_1^* \mathcal{L}_A) \stackrel{\mathrm{tr}(\pi_1)}{\to} H^i(X_\Delta(Q), \mathcal{L}_A).$$

This is the same definition as in [31, p. 586] and recovers the usual definition when $i = 0$. We define an action of T_x on the cohomology of $\mathcal{L} = \omega^n(-\infty)$, in a similar fashion: the operator xT_x is the composition

$$\begin{aligned} H^i(X_\Delta(Q), \omega^{\otimes n}(-\infty)_A) &\to H^i(X_\Delta(Q,x), (\pi_2^* \omega^{\otimes n})(-\infty)_A) \\ &\stackrel{\phi_{12}^{\otimes n}}{\to} H^i(X_\Delta(Q;x), (\pi_1^* \omega^{\otimes n})(-\infty)_A) \\ &\stackrel{\mathrm{tr}(\pi_1)}{\to} H^i(X_\Delta(Q), \omega^{\otimes n}(-\infty)_A). \end{aligned}$$

(In the first map, we use that $\pi_2^*(\omega^{\otimes n}(-\infty)) \subset (\pi_2^* \omega^{\otimes n})(-\infty)$ and in the last map we use that the trace maps sections which vanish at the cusps to sections which vanish at the cusps). Let $\mathcal{C}_\Delta(Q) \subset X_\Delta(Q)$ be the divisor of cusps, as above, and define $\mathcal{C}_\Delta(Q;x) \subset X_\Delta(Q;x)$ similarly. Then we define an action of T_x on $H^0(\mathcal{C}_\Delta(Q), \omega_A^n)$ by setting xT_x equal to the composition:

$$H^0(\mathcal{C}_\Delta(Q), \omega_A^{\otimes n}) \to H^0(\mathcal{C}_\Delta(Q,x), \pi_2^* \omega_A^{\otimes n}) \stackrel{\phi_{12}^{\otimes n}}{\to} H^0(\mathcal{C}_\Delta(Q;x), \pi_1^* \omega_A^{\otimes n}) \stackrel{\mathrm{tr}(\pi_1)}{\to} H^0(\mathcal{C}_\Delta(Q), \omega_A^{\otimes n}).$$

(In this line, both ϕ_{12} and the trace map are obtained from the previous ones by passing to the appropriate quotients). In this way the long exact sequence

$$\cdots \longrightarrow H^i(X_\Delta(Q), \omega^{\otimes n}(-\infty)_A) \longrightarrow H^i(X_\Delta(Q), \omega_A^{\otimes n}) \\ \longrightarrow H^i(\mathcal{C}_\Delta(Q), \omega_A^{\otimes n}) \to \cdots$$

is T_x-equivariant.

Remark 3.4 We note that the action of the operators T_x on $H^0(\mathcal{C}_\Delta(Q), \omega_A^{\otimes n})$ is given by specializing q equal to 0 in the usual q-expansion formula for the action of T_x (as in [28, Sect. 3], for instance). This expansion is with respect to a local parameter q at the cusp '∞', but note that the group $(\mathbf{Z}/NQ\mathbf{Z})^\times$ acts transitively on the set of cusps in $\mathcal{C}_\Delta(Q)$. In particular, suppose $f \in H^0(\mathcal{C}_\Delta(Q), \omega_k^{\otimes n})$ is a non-zero mod p eigenform for all T_x (with x prime to pNQ) and is of character $\chi : (\mathbf{Z}/NQ\mathbf{Z})^\times \to k^\times$ in the sense that $\langle a \rangle f = \chi(a) f$ for all $a \in (\mathbf{Z}/NQ\mathbf{Z})^\times$. Then f cannot vanish at any cusp and the formula [28, (3.5)] implies that for each x prime to pNQ, we have $T_x(f) = (1 + \chi(x)x^{n-1})f$. Thus, the semisimple Galois representation naturally associated to f (in the sense that for $x \nmid pNQ$, the representation is unramified at x with characteristic polynomial $X^2 - T_x X + \langle x \rangle x^{n-1}$) is $1 \oplus \chi\epsilon^{n-1}$ (where we also think of χ as a character of $G_{\mathbf{Q}}$ via class field theory).

For x a prime dividing Q, the action of U_x on $H^i(X_\Delta(Q), \mathcal{L}_A)$ (for $\mathcal{L} = \omega^{\otimes n}$, or $\omega^{\otimes n}(-\infty)$) and on $H^0(\mathcal{C}_\Delta(Q), \omega_A^{\otimes n})$ is defined similarly. In this case, we let $X_1(NQ; x)$ denote the smooth $\mathcal{O}$-curve parametrizing tuples (E, α_{NQ}, C_x) where E is a generalized elliptic curve, $\alpha_{NQ} : \mu_{NQ} \hookrightarrow E[NQ]$ is an embedding and $C_x \subset E[x]$ is a subgroup étale locally isomorphic to $\mathbf{Z}/x\mathbf{Z}$ such that $\alpha_{NQ}(\mu_{NQ}) + C_x$ meets every irreducible component in every geometric fiber and $\alpha_{NQ}(\mu_x) + C_x = E^{\mathrm{sm}}[x]$. We then let $X_\Delta(Q; x)$ be the quotient of $X_1(NQ; x)$ by the same subgroup used to define $X_\Delta(Q)$ as a quotient of $X_1(NQ)$. We have two projection maps $\pi_i : X_\Delta(Q; x) \to X_\Delta(Q)$ where π_1 is the forgetful map and π_2 sends (E, α_{NQ}, C_x) to $(E' = E/C_x, \alpha'_{NQ})$, as above. By [30, Proposition 4.4.3], we have a map $\phi_{12} : \pi_2^*\omega \to \pi_1^*\omega$ which allow us to define U_x by exactly the same formulas as above.

3.2.4 Properties of cohomology groups

Define the Hecke algebra

$$\mathbf{T}^{\mathrm{an}} \subset \mathrm{End}_{\mathcal{O}} H^0(X_1(Q), \omega_{K/\mathcal{O}})$$

to be the subring of endomorphisms generated over $\mathcal{O}$ by the Hecke operators T_n with $(n, pNQ) = 1$ together with all the diamond operators $\langle a \rangle$ for

$(a, NQ) = 1$. (Here "an" denotes anaemic). Let $\mathbf{T}$ denote the $\mathcal{O}$-algebra generated by these same operators together with U_x for x dividing Q. If $Q = 1$, we let $\mathbf{T}_\emptyset = \mathbf{T}^{\mathrm{an}}_\emptyset$ denote $\mathbf{T}$. The ring $\mathbf{T}^{\mathrm{an}}$ is a finite $\mathcal{O}$-algebra and hence decomposes as a direct product over its maximal ideals $\mathbf{T}^{\mathrm{an}} = \prod_{\mathfrak{m}} \mathbf{T}^{\mathrm{an}}_{\mathfrak{m}}$. We have natural homomorphisms

$$\mathbf{T}^{\mathrm{an}} \to \mathbf{T}^{\mathrm{an}}_\emptyset = \mathbf{T}_\emptyset, \qquad \mathbf{T}^{\mathrm{an}} \hookrightarrow \mathbf{T}$$

where the first is induced by the map $H^0(X, \omega_{K/\mathcal{O}}) \hookrightarrow H^0(X_1(Q), \omega_{K/\mathcal{O}})$ and the second is the obvious inclusion.

For each maximal ideal $\mathfrak{m}_\emptyset$ of $\mathbf{T}_\emptyset$, there is a finite extension k' of $\mathbf{T}_\emptyset/\mathfrak{m}_\emptyset$ and a continuous semisimple representation $G_{\mathbf{Q}} \to \mathrm{GL}_2(k')$ characterized by the fact that for each prime $x \nmid Np$, the representation is unramified at x and Frob_x has characteristic polynomial $X^2 - T_x X + \langle x \rangle$. (To see this, choose an extension k' and a normalized eigenform $f \in H^0(X, \omega_{k'})$ such that the action of $\mathbf{T}_\emptyset$ on f factors through $\mathfrak{m}_\emptyset$. Then apply [28, Proposition 11.1]). The representation may in fact be defined over the residue field $\mathbf{T}_\emptyset/\mathfrak{m}_\emptyset$. The maximal ideal $\mathfrak{m}_\emptyset$ is said to be *Eisenstein* if the associated Galois representation is reducible and *non-Eisenstein* otherwise. If $\mathfrak{m}_\emptyset$ is non-Eisenstein, then there is a continuous representation $G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{T}_{\emptyset,\mathfrak{m}_\emptyset})$ which is unramified away from pN and characterized by the same condition on characteristic polynomials.

Let $\mathfrak{m}_\emptyset$ be a non-Eisenstein maximal ideal of $\mathbf{T}_\emptyset$. By a slight abuse of notation, we also denote the preimage of $\mathfrak{m}_\emptyset$ in $\mathbf{T}^{\mathrm{an}}$ by $\mathfrak{m}_\emptyset$. Note that the resulting ideal $\mathfrak{m}_\emptyset \subset \mathbf{T}^{\mathrm{an}}$ is maximal. The localization $\mathbf{T}_{\mathfrak{m}_\emptyset}$ is a direct factor of $\mathbf{T}$ whose maximal ideals correspond (after possibly extending $\mathcal{O}$) to the U_x-eigenvalues on $H^0(X_1(Q), \omega_k)[\mathfrak{m}_\emptyset]$. There is a continuous representation $G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{T}^{\mathrm{an}}_{\mathfrak{m}_\emptyset})$ which is unramified away from pNQ and satisfies the same condition on characteristic polynomials as above for $x \nmid pNQ$. The following lemma is essentially well known in the construction of Taylor–Wiles systems, we give a detailed proof just to show that the usual arguments apply equally well in weight one.

Lemma 3.5 *Suppose that for each $x|Q$ we have that $x \equiv 1 \mod p$ and that the polynomial $X^2 - T_x X + \langle x \rangle \in \mathbf{T}_\emptyset[X]$ has distinct eigenvalues modulo $\mathfrak{m}_\emptyset$. Let $\mathfrak{m}$ denote the maximal ideal of $\mathbf{T}$ containing $\mathfrak{m}_\emptyset$ and $U_x - \alpha_x$ for some choice of root α_x of $X^2 - T_x X + \langle x \rangle \mod \mathfrak{m}_\emptyset$ for each $x|Q$. Then there is a $\mathbf{T}^{\mathrm{an}}_{\mathfrak{m}_\emptyset}$-isomorphism*

$$H^0(X, \omega_{K/\mathcal{O}})_{\mathfrak{m}_\emptyset} \xrightarrow{\sim} H^0(X_0(Q), \omega_{K/\mathcal{O}})_{\mathfrak{m}}.$$

Proof We first of all remark that the localization $H^0(X, \omega_{K/\mathcal{O}})_{\mathfrak{m}_\emptyset}$ is independent of whether we consider $\mathfrak{m}_\emptyset$ as an ideal of $\mathbf{T}_\emptyset$ or $\mathbf{T}^{\mathrm{an}}$. To see this, it suffices to note that $\mathbf{T}^{\mathrm{an}}_{\mathfrak{m}_\emptyset} \to \mathbf{T}_{\emptyset,\mathfrak{m}_\emptyset}$ is surjective. This in turn follows from the fact that the Galois representation $G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{T}_{\emptyset,\mathfrak{m}_\emptyset})$ can be defined over the image of $\mathbf{T}^{\mathrm{an}}_{\mathfrak{m}_\emptyset} \to \mathbf{T}_{\emptyset,\mathfrak{m}_\emptyset}$, and moreover that for $x|Q$, the operators T_x and $\langle x\rangle$ are given by the trace and determinant of Frob_x.

By induction, we reduce immediately to the case when $Q = x$ is prime. Let $\pi_1, \pi_2 : X_0(x) \to X$ denote the natural projection maps and let $\phi_{12} : \pi_2^*\omega \to \pi_1^*\omega$ be the map described in Sect. 3.2.3. We define $\psi := (\pi_1^*, \phi_{12} \circ \pi_2^*)$ and $\psi^\vee := \frac{1}{x}\binom{\mathrm{tr}(\pi_1)}{\mathrm{tr}(\pi_1)\circ W_x}$ where W_x is the operator defined in Sect. 3.2.3 (with NQ there playing the role of N here). These give a sequence of $\mathbf{T}^{\mathrm{an}}$-linear morphisms

$$H^0(X, \omega_{K/\mathcal{O}})^2 \xrightarrow{\psi} H^0(X_0(x), \omega_{K/\mathcal{O}}) \xrightarrow{\psi^\vee} H^0(X, \omega_{K/\mathcal{O}})^2,$$

such that the composite map $\psi^\vee \circ \psi$ is given by

$$\begin{pmatrix} x^{-1}\,\mathrm{tr}(\pi_1)\circ\pi_1^* & x^{-1}\,\mathrm{tr}(\pi_1)\circ\phi_{12}\circ\pi_2^* \\ x^{-1}\,\mathrm{tr}(\pi_1)\circ W_x\circ\pi_1^* & x^{-1}\,\mathrm{tr}(\pi_1)\circ W_x\circ\phi_{12}\circ\pi_2^* \end{pmatrix}$$
$$= \begin{pmatrix} x^{-1}(x+1) & T_x \\ T_x & \langle x\rangle(x+1) \end{pmatrix}.$$

(On the first row, this follows from the definition of T_x and the fact that π_1 has degree $x+1$; in the lower left corner we use the definition of T_x and the fact that $W_x \circ \pi_1^* = \phi_{12}\circ w_x^* \circ \pi_1^* = \phi_{12}\circ\pi_2^*$; in the lower right corner we use the facts that $W_x \circ \phi_{12} \circ \pi_2^* = W_x^2 \circ \pi_1^*$ and $W_x^2 = x\langle x\rangle$).

If α_x and β_x are the roots of $X^2 - T_x X + \langle x\rangle \mod \mathfrak{m}_\emptyset$, then $T_x \equiv \alpha_x + \beta_x \mod \mathfrak{m}_\emptyset$ and $\langle x\rangle \equiv \alpha_x\beta_x \mod \mathfrak{m}_\emptyset$. Since $x \equiv 1 \mod p$, we have

$$\begin{aligned}\det(\psi^\vee\circ\psi) &= x^{-1}(x+1)^2\langle x\rangle - T_x^2 \equiv 4\langle x\rangle - T_x^2 \\ &\equiv 4\alpha_x\beta_x - (\alpha_x+\beta_x)^2 \equiv -(\alpha_x-\beta_x)^2 \mod \mathfrak{m}_\emptyset.\end{aligned}$$

By assumption, $\alpha_x \not\equiv \beta_x$, and thus, after localizing at $\mathfrak{m}_\emptyset$, the composite map $\psi^\vee\circ\psi$ is an isomorphism.

We deduce that $H^0(X, \omega_{K/\mathcal{O}})^2_{\mathfrak{m}_\emptyset}$ is a direct factor of $H^0(X_0(x), \omega_{K/\mathcal{O}})_{\mathfrak{m}_\emptyset}$ as a $\mathbf{T}^{\mathrm{an}}_{\mathfrak{m}_\emptyset}$-module. Consider the action of U_x on the image of $H^0(X, \omega_{K/\mathcal{O}})^2$. Let V_x denote the second component $\phi_{12}\circ\pi_2^*$ of the degeneracy map ψ. Then we have equalities of maps $H^0(X, \omega_{K/\mathcal{O}}) \to H^0(X_0(x), \omega_{K/\mathcal{O}})$

$$\pi_1^*\circ T_x = U_x\circ\pi_1^* + \frac{1}{x}V_x \quad\text{and}\quad U_x\circ V_x = \pi_1^*\circ\langle x\rangle.$$

To see that the first of these holds, note that:

$$(\pi_1^* \circ T_x)(f)(E, \alpha_N, C_x) = \frac{1}{x} \sum_{D \subset E[x]} \phi_D^* f(E/D, \phi_D \circ \alpha_N),$$

where the sum is over all order x subgroups $D \subset E[x]$, and ϕ_D denotes the quotient map $E \to E/D$. (This formula holds after base-change to an étale extension over which the subgroups D are defined). Restricting the sum to all $D \neq C_x$ gives $(U_x \circ \pi_1^*)(f)(E, \alpha_N, C_x)$, while the remaining term is $\frac{1}{x}(\phi_{12} \circ \pi_2^*)(f)(E, \alpha_N, C_x)$. For the second equality, we have:

$$(U_x \circ V_x)(f)(E, \alpha_N, C_x) = \frac{1}{x} \sum_{D \neq C_x} \phi_{D+C_x}^* f(E/(D + C_x), \phi_{D+C_x} \circ \alpha_N),$$

where D is as above and ϕ_{D+C_x} is the quotient $E \to E/(D + C_x)$. Since $D+C_x = E[x]$ for all $D \neq C_x$, each term $\phi_{D+C_x}^* f(E/(D+C_x), \phi_{D+C_x} \circ \alpha_N)$ can be identified with $xf(E, x \circ \alpha_N) = x\langle x\rangle f(E, \alpha_N)$. Since there exactly x subgroups $D \neq C_x$, and we divide by x, we deduce that $U_x \circ V_x = \pi_1^*\langle x\rangle$.

It follows that the action of U_x on $H^0(X_0(x), \omega_{K/\mathcal{O}})^2_{\mathfrak{m}_\emptyset}$ is given by the matrix

$$A = \begin{pmatrix} T_x & x\langle x\rangle \\ -\frac{1}{x} & 0 \end{pmatrix}.$$

There is an identity $(A - \alpha_x)(A - \beta_x) \equiv 0 \mod \mathfrak{m}_\emptyset$ in $M_2(\mathbf{T}_{\emptyset,\mathfrak{m}_\emptyset})$. Since $\alpha_x \not\equiv \beta_x$, by Hensel's Lemma, there exist $\widetilde{\alpha}_x$ and $\widetilde{\beta}_x$ in $\mathbf{T}^\times_{\emptyset,\mathfrak{m}_\emptyset}$ such that $(U_x - \widetilde{\alpha}_x)(U_x - \widetilde{\beta}_x) = 0$ on $(\mathrm{Im}\,\psi)_{\mathfrak{m}_\emptyset}$. It follows that $U_x - \widetilde{\beta}_x$ is a projector (up to a unit) from $(\mathrm{Im}\,\psi)_{\mathfrak{m}_\emptyset}$ to $(\mathrm{Im}\,\psi)_{\mathfrak{m}}$. We claim that there is an isomorphism of $\mathbf{T}^{\mathrm{an}}_{\mathfrak{m}_\emptyset}$-modules

$$H^0(X, \omega_{K/\mathcal{O}})_{\mathfrak{m}_\emptyset} \simeq (\mathrm{Im}\,\psi)_{\mathfrak{m}} = (H^0(X, \omega_{K/\mathcal{O}})^2)_{\mathfrak{m}}.$$

It suffices to show that there is a $\mathbf{T}^{\mathrm{an}}_{\mathfrak{m}_\emptyset}$-equivariant injection from $H^0(X, \omega_{K/\mathcal{O}})_{\mathfrak{m}_\emptyset}$ to the module $(\mathrm{Im}\,\psi)_{\mathfrak{m}_\emptyset}$ such that the image has trivial intersection with the kernel of $U_x - \widetilde{\beta}_x$: if there is such an injection, then, by symmetry, there is also an injection from $H^0(X, \omega_{K/\mathcal{O}})_{\mathfrak{m}_\emptyset}$ to $(\mathrm{Im}\,\psi)_{\mathfrak{m}_\emptyset}$ whose image intersects the kernel of $U_x - \widetilde{\alpha}_x$ trivially; by length considerations both injections are forced to be isomorphisms. We claim that the natural inclusion π_1^* composed with $U_x - \widetilde{\beta}_x$ is such a map: from the computation of the matrix above, it follows that

$$(U_x - \widetilde{\beta}_x)\begin{pmatrix} f \\ 0 \end{pmatrix} = \begin{pmatrix} T_x f - \widetilde{\beta}_x f \\ -f \end{pmatrix},$$

which is non-zero whenever f is by examining the second coordinate.

We deduce that there is a decomposition of $\mathbf{T}^{\mathrm{an}}_{\mathfrak{m}_\emptyset}$-modules

$$H^0(X_0(x), \omega_{K/\mathcal{O}})_{\mathfrak{m}} = H^0(X, \omega_{K/\mathcal{O}})_{\mathfrak{m}_\emptyset} \oplus V$$

where V is the kernel of $\psi^\vee$. It suffices to show that $V[\mathfrak{m}]$ is zero. Let $f \in V[\mathfrak{m}]$. We may regard f as an element of $H^0(X_0(x), \omega_k)$. It satisfies the following properties:

(1) $U_x f = \alpha_x f$,
(2) $\langle x \rangle f = \alpha_x \beta_x f$,
(3) $(\mathrm{tr}(\pi_1) \circ W_x) f = 0$.

The first two properties follow from the fact that f is killed by $\mathfrak{m}$, and the last follows from the fact that f lies in the kernel of $\psi^\vee$.

We claim that

$$\frac{1}{x}(\pi_1^* \circ \mathrm{tr}(\pi_1) \circ W_x) = U_x + \frac{1}{x} W_x.$$

To see this, we will rewrite the relation $\pi_1^* \circ T_x = U_x \circ \pi_1^* + \frac{1}{x} V_x$ that we established earlier. Since $\pi_2 = \pi_1 \circ w_x$, we have that $T_x = \frac{1}{x}\,\mathrm{tr}(\pi_1) \circ W_x \circ \pi_1^*$ and $V_x = W_x \circ \pi_1^*$. The earlier relation can thus be written:

$$\frac{1}{x}\pi_1^* \circ \mathrm{tr}(\pi_1) \circ W_x \circ \pi_1^* = U_x \circ \pi_1^* + \frac{1}{x} W_x \circ \pi_1^*.$$

Since π_1^* is fully faithful, the claim follows.

Now, property (3) above tells us that $-xU_x f = W_x f$. By (1), both sides are k-multiples of f and applying $-xU_x f = W_x f$ once more, we see that $x^2 U_x^2 f = W_x^2 f$. Since $W_x^2 = x\langle x \rangle$, we deduce that $x U_x^2 f = \langle x \rangle f$. Thus, by (1) and (2):

$$x \alpha_x^2 f = \alpha_x \beta_x f.$$

Since $x \equiv 1 \mod p$ and $\alpha_x \neq \beta_x$, we deduce that $f = 0$, as required. □

We have the following mild generalization of Lemma 3.5:

Lemma 3.6 *Suppose $Q = x$ is prime not dividing Np and the eigenvalues of the polynomial $X^2 - T_x X + \langle x \rangle \in \mathbf{T}_\emptyset/\mathfrak{m}_\emptyset[X]$ do not have ratio x or x^{-1}. Let $\mathfrak{m}$ denote the maximal ideal of $\mathbf{T}$ containing $\mathfrak{m}_\emptyset$ and $U_x - \alpha_x$ for some choice* *of root α_x of $X^2 - T_x X + \langle x \rangle \mod \mathfrak{m}_\emptyset$. If the roots α_x and β_x are distinct, then there is a $\mathbf{T}^{\mathrm{an}}_{\mathfrak{m}_\emptyset}$-isomorphism*

$$H^0(X, \omega_{K/\mathcal{O}})_{\mathfrak{m}_\emptyset} \xrightarrow{\sim} H^0(X_0(x), \omega_{K/\mathcal{O}})_{\mathfrak{m}}.$$

If the roots α_x and β_x are equal, then there is a $\mathbf{T}^{\mathrm{an}}_{\mathfrak{m}_\emptyset}$-isomorphism

$$H^0(X, \omega_{K/\mathcal{O}})^2_{\mathfrak{m}_\emptyset} \xrightarrow{\sim} H^0(X_0(x), \omega_{K/\mathcal{O}})_{\mathfrak{m}}.$$

Proof The proof is essentially identical to Lemma 3.5. The only calculations which are different are the following: If α_x and β_x are the roots of $X^2 - T_x X + \langle x \rangle \mod \mathfrak{m}_\emptyset$, then $T_x \equiv \alpha_x + \beta_x \mod \mathfrak{m}_\emptyset$ and $\langle x \rangle \equiv \alpha_x \beta_x \mod \mathfrak{m}_\emptyset$. Hence

$$\begin{aligned}\det(\psi^\vee \circ \psi) &= x^{-1}(x+1)^2 \langle x \rangle - T_x^2 \\ &\equiv x^{-1}((x+1)^2 \alpha_x \beta_x - x(\alpha_x + \beta_x)^2) \\ &\equiv x^{-1}(\alpha_x - x\beta_x)(\beta_x - x\alpha_x) \mod \mathfrak{m}_\emptyset.\end{aligned}$$

This is non-zero under our assumptions. If the eigenvalues are distinct, we proceed as before; the proof that the summand V of $H^0(X_0(Q), \omega_{K/\mathcal{O}})_{\mathfrak{m}}$ is zero is also the same, since the final conclusion is that $\beta_x \equiv x\alpha_x \mod \mathfrak{m}_\emptyset$, a contradiction. If the eigenvalues are equal, note that

$$\psi(H^0(X, \omega_{K/\mathcal{O}})^2_{\mathfrak{m}_\emptyset}) = \psi(H^0(X, \omega_{K/\mathcal{O}})^2_{\mathfrak{m}_\emptyset})_{\mathfrak{m}}$$

since $\mathbf{T}$ acts on this space via the quotient $\mathbf{T}_{\emptyset, \mathfrak{m}_\emptyset}[X]/(X^2 - T_x X + \langle x \rangle)$ (with X corresponding to U_x), which is a local ring, by assumption. Thus, we can decompose

$$H^0(X_0(x), \omega_{K/\mathcal{O}})_{\mathfrak{m}} = \psi(H^0(X, \omega_{K/\mathcal{O}})_{\mathfrak{m}_\emptyset})^2 \oplus V,$$

as a direct sum of $\mathbf{T}^{\mathrm{an}}$-modules. The same proof as above shows that V is zero. □

As in Sect. 3.2.3, let Δ be a quotient of $\Delta_Q := (\mathbf{Z}/Q\mathbf{Z})^\times$ and let $X_\Delta(Q) \to X_0(Q)$ be the corresponding cover. If A is an $\mathcal{O}$-module, we have defined an action of the universal polynomial algebra $\mathbf{T}^{\mathrm{univ}}$ on the cohomology groups $H^i(X_\Delta(Q), \mathcal{L}_A)$ for $\mathcal{L} = \omega^{\otimes n}$ or $\omega^{\otimes n}(-\infty)$. The ideal $\mathfrak{m}_\emptyset$ gives rise to a maximal ideal $\mathfrak{m}$ of $\mathbf{T}^{\mathrm{univ}}$ after a choice of eigenvalue mod $\mathfrak{m}_\emptyset$ for U_x for all x dividing Q. Extending $\mathcal{O}$ if necessary, we may assume that $\mathbf{T}^{\mathrm{univ}}/\mathfrak{m} \cong k$.

Let $M \mapsto M^\vee := \mathrm{Hom}_{\mathcal{O}}(M, K/\mathcal{O})$ denote the Pontryagin duality functor.

Lemma 3.7 *Let Δ be a quotient of Δ_Q. Then:*

(1) $H^1(X_\Delta(Q), \mathcal{L}_{K/\mathcal{O}})^\vee$ *is p-torsion free for* $\mathcal{L}$ *a vector bundle on* $X_\Delta(Q)$.
(2) *For* $i = 0, 1$, *we have an isomorphism*

$$H^i(X_\Delta(Q), \omega(-\infty)_{K/\mathcal{O}})_\mathfrak{m} \xrightarrow{\sim} H^i(X_\Delta(Q), \omega_{K/\mathcal{O}})_\mathfrak{m}.$$

Proof The first claim is equivalent to the divisibility of $H^1(X_\Delta(Q), \mathcal{L}_{K/\mathcal{O}})$. Since $X_\Delta(Q)$ is flat over $\mathcal{O}$, there is an exact sequence

$$0 \to \mathcal{L}_k \to \mathcal{L}_{K/\mathcal{O}} \xrightarrow{\varpi} \mathcal{L}_{K/\mathcal{O}} \to 0.$$

Taking cohomology and using the fact that $X_\Delta(Q)$ is a curve and hence $H^2(X_\Delta(Q), \mathcal{L}_k)$ vanishes, we deduce that $H^1(X_\Delta(Q), \mathcal{L}_{K/\mathcal{O}})/\varpi = 0$, from which divisibility follows.

For the second claim, note that there is an exact sequence:

$$0 \to \omega(-\infty) \to \omega \to \omega|_{\mathcal{C}_\Delta(Q)} \to 0,$$

where $\mathcal{C}_\Delta(Q)$ denote the divisor of cusps. The cohomology of $\mathcal{C}_\Delta(Q)$ is concentrated in degree 0. Yet the action of Hecke on $H^0(\mathcal{C}_\Delta(Q), \omega)$ is Eisenstein (by Remark 3.4), thus the lemma. □

If $\mathcal{L}$ is a vector bundle on $X_\Delta(Q)$, we define

$$H_i(X_\Delta(Q), \mathcal{L}) := H^i(X_\Delta(Q), (\Omega^1 \otimes \mathcal{L}^*)_{K/\mathcal{O}})^\vee$$

for $i = 0, 1$, where $\mathcal{L}^*$ is the dual bundle and $\Omega^1 = \Omega^1_{X_\Delta(Q)/\mathcal{O}}$ can be identified with $\omega^{\otimes 2}(-\infty)$ via the Kodaira–Spencer isomorphism. If $\Delta \twoheadrightarrow \Delta'$ are two quotients of $(\mathbf{Z}/Q\mathbf{Z})^\times$ giving rise to a Galois cover $\pi : X_\Delta(Q) \to X_{\Delta'}(Q)$ and $\mathcal{L}$ is vector bundle on $X_{\Delta'}(Q)$, then there is a natural map $\pi_* : H_i(X_\Delta(Q), \pi^*\mathcal{L}) \to H_i(X_{\Delta'}(Q), \mathcal{L})$ coming from the dual of the pullback π^* on cohomology. Verdier duality [32, Corollary 11.2(f)] gives an isomorphism

$$D : H_i(X_\Delta(Q), \mathcal{L}) \xrightarrow{\sim} H^{1-i}(X_\Delta(Q), \mathcal{L})$$

under which π_* corresponds to the trace map $\mathrm{tr}(\pi) : H^{1-i}(X_\Delta(Q), \pi^*\mathcal{L}) \to H^{1-i}(X_{\Delta'}(Q), \mathcal{L})$.

We endow $H_i(X_\Delta(Q), \omega^{\otimes n})$ with a Hecke action by first identifying Ω^1 with $\omega^2(-\infty)$ and then taking the Pontryagin dual of the Hecke action on $H^i(X_\Delta(Q), \omega^{2-n}(-\infty)_{K/\mathcal{O}})$. However, we note that for $\mathcal{L} = \omega^n$, the isomorphism D is not Hecke equivariant: extend $\mathcal{O}$ if necessary so that it contains a primitive NQ-th root of unity ζ. Let w^* be the operator on $H^0(X_\Delta(Q), \omega^{\otimes n}_{K/\mathcal{O}})$

associated to ζ as in [31, Sect. 7.1]. Let Φ denote the composition of isomorphisms:

$$\begin{aligned} H^1(X_\Delta(Q), \omega^{2-n}(-\infty)_{K/\mathcal{O}}) &\xrightarrow{D} H^0(X_\Delta(Q), \Omega \otimes \omega^{n-2}(\infty))^\vee \\ &\xrightarrow{KS^\vee} H^0(X_\Delta(Q), \omega^n)^\vee \\ &\xrightarrow{(w^*)^\vee} H^0(X_\Delta(Q), \omega^n)^\vee, \end{aligned}$$

where D is Verdier duality, and KS is the Kodaira–Spencer isomorphism. Then by the proof of [31, Proposition 7.3], we have:

$$\Phi \circ T_x = x^{1-n} T_x^\vee \circ \Phi,$$

with the same relation holding for U_x. We also have $\Phi \circ \langle x \rangle = \langle x \rangle \circ \Phi$. We let $\Psi := \Phi^\vee$ be the dual isomorphism

$$\Psi : H^0(X_\Delta(Q), \omega^n) \longrightarrow H_1(X_\Delta(Q), \omega^n).$$

When $n = 1$, Ψ is $\mathbf{T}$-linear.

Proposition 3.8 *Let Δ be a quotient of $(\mathbf{Z}/Q\mathbf{Z})^\times$ of p-power order. Then the $\mathcal{O}[\Delta]$-module $H_0(X_\Delta(Q), \omega)_\mathfrak{m}$ is balanced (in the sense of Definition 2.2).*

Proof Let $M = H_0(X_\Delta(Q), \omega \otimes \mathcal{L})_\mathfrak{m}$ and $S = \mathcal{O}[\Delta]$, where $\mathcal{L} = \mathcal{O}_X$.[2] Consider the exact sequence of S-modules (with trivial Δ-action):

$$0 \to \mathcal{O} \xrightarrow{\varpi} \mathcal{O} \to k \to 0.$$

Tensoring this exact sequence over S with M, we obtain an exact sequence:

$$0 \to \operatorname{Tor}_1^S(M, \mathcal{O})/\varpi \to \operatorname{Tor}_1^S(M, k) \to M_\Delta \to M_\Delta \to M \otimes_S k \to 0.$$

Let r denote the $\mathcal{O}$-rank of M_Δ. Then this exact sequence tells us that

$$d_S(M) := \dim_k M \otimes_S k - \dim_k \operatorname{Tor}_1^S(M, k) = r - \dim_k \operatorname{Tor}_1^S(M, \mathcal{O})/\varpi.$$

We have a second quadrant Hochschild–Serre spectral sequence [33, Theorem III.2.20, Remark III.3.8]

$$\begin{aligned} &H^i(\Delta, H^j(X_\Delta(Q), (\Omega^1 \otimes \omega^{-1} \otimes \mathcal{L}^*)_{K/\mathcal{O}})) \\ &\quad \Longrightarrow H^{i+j}(X_0(Q), (\Omega^1 \otimes \omega^{-1} \otimes \mathcal{L}^*)_{K/\mathcal{O}}). \end{aligned}$$

[2] We present the proof writing $\mathcal{L}$ instead of $\mathcal{O}_X$ since we will use the same argument in the proof of Theorem 3.30 with $\mathcal{L}$ as a more general vector bundle on X.

Applying Pontryagin duality, we obtain a third quadrant spectral sequence

$$H_i(\Delta, H_j(X_\Delta(Q), \omega \otimes \mathcal{L})) = \mathrm{Tor}_i^S(H_j(X_\Delta(Q), \omega \otimes \mathcal{L}), \mathcal{O}) \\ \implies H_{i+j}(X_0(Q), \omega \otimes \mathcal{L}).$$

We claim that the differentials in the spectral sequence commute with the action of $\mathbf{T}^{\mathrm{univ}}$ on the individual terms. This follows from the fact that the Hochschild–Serre spectral sequence, for a finite étale Galois cover $\pi : X \to Y$ with group G,

$$H^i(G, H^j(X, \pi^*\mathcal{F})) \implies H^{i+j}(Y, \mathcal{F})$$

is functorial in $\mathcal{F}$. Thus for example, to see that the differentials commute with

$$T_x = \frac{1}{x}\,\mathrm{tr}(\pi_1) \circ \phi_{12} \circ \pi_2^*,$$

(where $\pi_1, \pi_2 : X_*(Q; x) \to X_*(Q)$ are the two projection maps and $* \in \{0, \Delta\}$), we use the canonical isomorphisms $H^i(X_*(Q; x), \pi_j^*\omega) = H^i(X_*(Q), \pi_{j,*}\pi_j^*\omega)$ and successively apply the functoriality of the spectral sequence with $(X, Y, \mathcal{F} \to \mathcal{F}')$ taken to equal $(X_\Delta(Q), X_0(Q), \omega \to \pi_{2,*}\pi_2^*\omega)$, $(X_\Delta(Q; x), X_0(Q; x), \pi_2^*(\omega) \stackrel{\phi_{12}}{\to} \pi_1^*(\omega))$ and $(X_\Delta(Q), X_0(Q), \pi_{1,*}\pi_1^*(\omega) \stackrel{\mathrm{tr}}{\to} \omega)$.

Localizing at $\mathfrak{m}$, we obtain an isomorphism $M_\Delta \cong H_0(X_0(Q), \omega \otimes \mathcal{L})_\mathfrak{m}$ and an exact sequence

$$(H_1(X_\Delta(Q), \omega \otimes \mathcal{L})_\mathfrak{m})_\Delta \to H_1(X_0(Q), \omega \otimes \mathcal{L})_\mathfrak{m} \to \mathrm{Tor}_1^S(M, \mathcal{O}) \to 0.$$

To show that $d_S(M) \geq 0$, we see that it suffices to show that $H_1(X_0(Q), \omega \otimes \mathcal{L})_\mathfrak{m}$ is free of rank r as an $\mathcal{O}$-module. The module $H_1(X_0(Q), \omega \otimes \mathcal{L})$ is p-torsion free by Lemma 3.7 (1). It therefore suffices to show that $\dim_K H_1(X_0(Q), \omega \otimes \mathcal{L})_\mathfrak{m} \otimes K = r$. In other words, by the definition of r, it suffices to show that

$$\dim_K H_0(X_0(Q)_K, \omega \otimes \mathcal{L})_\mathfrak{m} = \dim_K H_1(X_0(Q), \omega \otimes \mathcal{L})_\mathfrak{m}.$$

(Here we use the slight abuse of notation $H^i(X_K, *)_\mathfrak{m} = H^i(X, *)_\mathfrak{m} \otimes_\mathcal{O} K$).

At this point we will specialize to $\mathcal{L} = \mathcal{O}_X$. By definition of H_0 and its Hecke action, the left hand side above is:

$$\dim_K H^0(X_0(Q)_K, \omega(-\infty))_\mathfrak{m}.$$

Using the isomorphism Ψ, we see that the right hand side is equal to:

$$H^0(X_0(Q)_K, \omega)_{\mathfrak{m}}$$

We are therefore reduced to showing that

$$\dim_K H^0(X_0(Q)_K, \omega(-\infty))_{\mathfrak{m}} = \dim_K H^0(X_0(Q)_K, \omega)_{\mathfrak{m}}.$$

The result thus follows from Lemma 3.7 (2). □

3.3 Galois representations

Let $N = N(\overline{\rho})$ where $N(\overline{\rho})$ is the Serre level of $\overline{\rho}$. We let H denote the p-part of $(\mathbf{Z}/N(\overline{\rho})\mathbf{Z})^\times$. Having fixed N and H, we let X denote the modular curve defined at the beginning of Sect. 3.2. (We note that $N \geq 5$ by Serre's conjecture, and also that $\overline{\rho}$ is thus modular of the appropriate level, by Theorem 4.5 of [31]). We add, for now, the following assumption:

Assumption 3.9 Assume that:

(1) The set $T(\overline{\rho})$ is empty.

We address how to remove this assumption in Sect. 3.9. The only point at which this assumption is employed is in Sect. 3.4.

Remark 3.10 **A digression on representability and stacks**. Several of the modular curves we consider do not represent the corresponding moduli problems for elliptic curves with level structure (due to automorphisms). In such cases, the object $X_H(N)$ still exists as a smooth proper Deligne–Mumford stack over $\mathrm{Spec}(\mathcal{O})$, and the sheaf ω descends to a sheaf on $X_H(N)$ (if not always to the corresponding associated scheme). For these stacks X, one can still make sense of the cohomology groups $H^0(X, \omega)$ and show that they satisfy many of the required properties. For example, suppose that $Y \to X$ is a finite étale morphism of modular stacks with Galois group G, and that $\mathscr{L} = \omega^k$ for some k. Then there is an isomorphism

$$H^0(X, \mathscr{L}) \simeq H^0(Y, \mathscr{L})^G.$$

Taking Y to be representable (which is always possible for the X we consider), and letting R be an $\mathcal{O}$-algebra, one may identify $H^0(X_R, \omega^k)$ with the ring of Katz modular forms of weight k over R as defined in [29].[3]

[3] Here is a somewhat different example: let $H \subset (\mathbf{Z}/13\mathbf{Z})^\times$ denote the group of squares; there is a finite étale map $X_1(13) \to X_H(13)$ with Galois group $\mathbf{Z}/3\mathbf{Z}$ (viewing $X_H(13)$ as a stack). The underlying scheme of $X_H(13)$ is isomorphic over $\mathcal{O}$ (for $p \neq 13$) to the projective

Instead of trying to adapt our arguments (when necessary) to the context of stacks, we introduce the following fix. Choose any prime $q \not\equiv 1 \mod p$ with $q \geq 5$ such that $\overline{\rho}$ is unramified at q and such that the ratio of the eigenvalues of $\overline{\rho}(\mathrm{Frob}_q)$ is neither q nor q^{-1} (we allow the possibility that the eigenvalues are the same). The assumption on the ratio of the eigenvalues ensures that $\overline{\rho}$ admits no deformations which are ramified and Steinberg (unipotent on inertia) at q; the assumption that $q \not\equiv 1 \mod p$ guarantees that there are no other deformations of $\overline{\rho}$ which are ramified at q. First, we claim that the Chebotarev density theorem guarantees the existence of such primes. Since $\det(\overline{\rho})$ is unramified at p, the fixed field of $\ker(\overline{\rho})$ does not contain $\mathbf{Q}(\zeta_p)$ (note that $p \neq 2$). Hence, we may find infinitely many primes q such that the fixed field of $\ker(\overline{\rho})$ splits completely and $q \not\equiv 1 \mod p$; such primes satisfy the required hypothesis. We then add $\Gamma_1(q)$ level structure and the arguments proceed almost entirely unchanged (the assumption on q implies that any deformation of $\overline{\rho}$ with fixed determinant is unramified at q). The only difference is that the multiplicity of the corresponding Hecke modules will either be the same or twice as expected—depending on whether $\overline{\rho}(\mathrm{Frob}_q)$ has distinct eigenvalues or not—by Lemma 3.6.

By Serre's conjecture [24] and by the companion form result of Gross [28] and Coleman–Voloch [37], there exists a maximal ideal $\mathfrak{m}_\emptyset$ of $\mathbf{T}_\emptyset$ corresponding to $\overline{\rho}$. The ideal $\mathfrak{m}_\emptyset$ is generated by ϖ, $T_x - \mathrm{Trace}(\overline{\rho}(\mathrm{Frob}_x))$ for all primes x with $(x, Np) = 1$ and $\langle x \rangle - \det(\overline{\rho}(\mathrm{Frob}_x))$ for all x with $(x, N) = 1$. Extending $\mathcal{O}$ if necessary, we may assume $\mathbf{T}_\emptyset/\mathfrak{m}_\emptyset = k$. Let Q be as in Sect. 3.2. For each $x \in Q$, assume that the polynomial $X^2 - T_x X + \langle x \rangle$ has distinct roots in $\mathbf{T}_\emptyset/\mathfrak{m}_\emptyset = k$ and choose a root $\alpha_x \in k$ of this polynomial. Let $\mathfrak{m}$ denote the maximal ideal of $\mathbf{T}$ generated by $\mathfrak{m}_\emptyset$ and $U_x - \alpha_x$ for $x \in Q$.

Theorem 3.11 (Local–Global Compatibility) *There exists a deformation*

$$\rho_Q : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{T}_\mathfrak{m})$$

of $\overline{\rho}$ *unramified outside* NQ *and determined by the property that for all primes* x *satisfying* $(x, pNQ) = 1$, $\mathrm{Trace}(\rho_Q(\mathrm{Frob}_x)) = T_x$. *Let* $\rho'_Q =$

Footnote 3 continued
line, and hence, naïvely, one would expect $H^0(X_H(13), \omega^2)$ to vanish. However, as noted by Serre [34,35], there exists a Galois representation $\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{F}_3)$ with $N(\overline{\rho}) = 13$, $k(\overline{\rho}) = 2$, and $\epsilon(\overline{\rho})$ quadratic. (The representation $\overline{\rho}$ is induced from $\mathbf{Q}(\sqrt{-3})$). The original conjecture 3.2.4? of [36] asserts that $\overline{\rho}$ gives rise to a mod-3 modular form on $X_H(13)$. However, considering $X_H(13)$ as a stack, one finds, for $p = 3$, that the group

$$H^0\left(X_H(13)_{\mathbf{F}_3}, \omega^2\right) = H^0\left(X_1(13)_{\mathbf{F}_3}, \omega^2\right)^{\mathbf{Z}/3\mathbf{Z}}$$

is indeed non-zero, even though the scheme underlying $X_H(13)$ has genus zero. This is in accordance with Edixhoven's reformulation of Serre's conjecture (Conjecture 4.2 of [31]).

$\rho_Q \otimes \eta$, where η is the unique p-power order character satisfying $\eta^2 = \langle \det(\overline{\rho}) \rangle \det(\rho_Q)^{-1}$. Then ρ'_Q is a deformation of $\overline{\rho}$ minimal outside Q. For a prime $x|Q$, the restriction of ρ'_Q to D_x is conjugate to a direct sum of residually unramified characters $\xi \oplus \langle \det(\overline{\rho}) \rangle |_{D_x} \xi^{-1}$ where $\overline{\xi}(\mathrm{Frob}_x) \equiv \alpha_x \mod \mathfrak{m}$, and such that the restriction of ξ to I_x and hence via local class field theory to $\mathbf{Z}_x^\times \otimes \mathbf{Z}_p = \Delta_x$ is compatible in the usual way with the diamond operators in $\mathbf{T}_\mathfrak{m}$.

Remark 3.12 The existence of ρ_Q follows immediately by using congruences between weight one forms and higher weight (using powers of the Hasse invariant). The assumptions on $x|Q$ imply that $\overline{\rho}|_{D_x}$ is p-distinguished and, because there are no local extensions, totally split. Hence the required information is preserved under congruences, and one is reduced once more to higher weight, where this statement is known. Hence the main difficulty in proving Theorem 3.11 is showing that the Galois representation is unramified at p.

Remark 3.13 Under the hypothesis that $\overline{\rho}(\mathrm{Frob}_p)$ has distinct eigenvalues, Theorem 3.11 can be deduced using an argument similar to that of [28]. Under the hypothesis that $\overline{\rho}(\mathrm{Frob}_p)$ has repeated eigenvalues but is *not* scalar, we shall deduce this using an argument of Wiese [38] and Buzzard. When $\overline{\rho}(\mathrm{Frob}_p)$ is trivial, however, we shall be forced to find a new argument using properties of local deformation rings. In the argument below, we avoid using the fact that the Hecke eigenvalues for all primes l determine a modular eigenform completely. One reason for doing this is that we would like to generalize our arguments to situations in which this fact is no longer true; we apologize in advance that this increases the difficulty of the argument slightly (specifically, we avoid using the fact that T_p in weight one can be shown to live inside the Hecke algebra $\mathbf{T}$, although this will be a consequence of our results).

Proof For each $m > 0$, we have $H^0(X_1(Q), \omega_{\mathcal{O}/\varpi^m})_\mathfrak{m} \cong H^0(X_1(Q), \omega_{K/\mathcal{O}})_\mathfrak{m}[\varpi^m]$, and we let I_m denote the annihilator of this space in $\mathbf{T}_\mathfrak{m}$. Since $\mathbf{T}_\mathfrak{m} = \varprojlim_m \mathbf{T}_\mathfrak{m}/I_m$, it suffices to construct, for each $m > 0$, a representation $\rho_{Q,m} : G_\mathbf{Q} \to \mathrm{GL}_2(\mathbf{T}_\mathfrak{m}/I_m)$ satisfying the conditions of the theorem.

Fix $m > 0$ and let A be a lift of (some power of) the Hasse invariant such that $A \equiv 1 \mod \varpi^m$; let $n-1$ denote the weight of A. We may assume that $n-1$ is sufficiently divisible by powers of p (and $(p-1)$) to ensure that $\epsilon^{n-1} \equiv 1 \mod \varpi^m$. Multiplication by A induces a map:

$$\begin{array}{ccc} H^0(X_1(Q), \omega_{K/\mathcal{O}}) & \xrightarrow[\phi]{} & H^0\left(X_1(Q), \omega^n_{K/\mathcal{O}}\right) \\ \downarrow & & \downarrow \\ K/\mathcal{O}[q] & = & K/\mathcal{O}[q] \end{array}.$$

This map is Hecke equivariant away from p on ϖ^m torsion; indeed, the diagram is only commutative modulo ϖ^m.

Consider the map

$$\begin{array}{ccc} \psi : H^0\left(X_1(Q), \omega_{K/\mathcal{O}}\right)^2[\varpi^m] & \longrightarrow & H^0\left(X_1(Q), \omega^n_{K/\mathcal{O}}\right)[\varpi^m] \\ \| & & \| \\ \psi : H^0\left(X_1(Q), \omega_{\mathcal{O}/\varpi^m}\right)^2 & \longrightarrow & H^0\left(X_1(Q), \omega^n_{\mathcal{O}/\varpi^m}\right). \end{array}$$

defined by $\psi = (\phi, \phi \circ T_p - U_p \circ \phi)$. (The operator T_p acts in this setting by the results of Sect. 4 of Gross [28]). For ease of notation, we let T_p (or $\phi \circ T_p$) exclusively refer to the Hecke operator in weight one, and let U_p denote the corresponding Hecke operator in weight n. (The operator U_p has the expected effect on q-expansions, since the weight n is sufficiently large with respect to m). On q-expansions modulo ϖ^m, we may compute that $\psi = (\phi, \langle p \rangle V_p)$ (see also 4.7 of [28]) . We claim that ψ is injective. It suffices to check this on the $\mathcal{O}$-socle, namely, on ϖ-torsion. On q-expansions, ϕ is the identity and $V_p(\sum a_n q^n) = \sum a_n q^{np}$. Suppose we have an identity $\langle p \rangle V_p f = g$. It follows that $\theta g = 0$ in $H^0(X_1(Q), \omega_{\mathcal{O}/\varpi^m})[\varpi] = H^0(X_1(Q), \omega_k)$. By a result of Katz [29], the θ map has no kernel in weight $\leq p-2$, and so in particular no kernel in weight 1. Hence ψ is injective.

The action of U_p in weight n on $H^0(X_1(Q), \omega_{\mathcal{O}/\varpi^m})^2$ via ψ^{-1} is given by

$$\begin{pmatrix} T_p & 1 \\ -\langle p \rangle & 0 \end{pmatrix},$$

where here T_p is acting in weight one (cf. Prop 4.1 of [28]), and hence satisfies the quadratic relation $X^2 - T_p X + \langle p \rangle = 0$. Note that the action of $U_p + \langle p \rangle U_p^{-1}$ on the image of ψ is given by

$$\begin{pmatrix} T_p & 0 \\ 0 & T_p \end{pmatrix}.$$

By Proposition 12.1 and the remark before equation (4.7) of [28], we see that $\langle p \rangle = \alpha\beta \mod \mathfrak{m}$ and $(U_p - \widetilde{\alpha})(U_p - \widetilde{\beta})$ acts nilpotently on $\psi(H^0(X_1(Q), \omega_{\mathcal{O}/\varpi^m})^2_{\mathfrak{m}})$, where α and β denote the (possibly non-distinct) eigenvalues of $\overline{\rho}(\mathrm{Frob}_p)$ and $\widetilde{\alpha}$, $\widetilde{\beta}$ are any lifts of α and β to $\mathcal{O}$. Explicitly, the only possible eigenvalues of U_p modulo $\mathfrak{m}$ in higher weight are determined by $\overline{\rho}$, and are either equal to α, β, or 0. Yet U_p acts invertibly on $\psi(H^0(X_1(Q), \omega_{\mathcal{O}/\varpi^m})^2_{\mathfrak{m}})$, as can be seen by considering the matrix description of U_p (which has invertible determinant given by $\langle p \rangle$).

Let $\mathbf{T}_n^{\mathrm{an}}$ denote the subalgebra of $\mathrm{End}_{\mathcal{O}}(H^0(X_1(Q), \omega_{\mathcal{O}}^n))$ generated over $\mathcal{O}$ by the operators T_x for primes $(x, NQp) = 1$ and diamond operators $\langle a \rangle$ for $(a, NQ) = 1$. Let $\mathbf{T}_n$ denote the subalgebra of $\mathrm{End}_{\mathcal{O}}(H^0(X_1(Q), \omega_{\mathcal{O}}^n))$ generated by $\mathbf{T}_n^{\mathrm{an}}$ and U_x for x dividing Q, and let $\widetilde{\mathbf{T}}_n$ denote the subalgebra generated by $\mathbf{T}_n$ and U_p (recall that we are denoting T_p in weight n by U_p). By a slight abuse of notation, let $\mathfrak{m}_\emptyset$ denote the maximal ideal of $\mathbf{T}_n^{\mathrm{an}}$ corresponding to $\overline{\rho}$. Similarly, let $\mathfrak{m}$ denote the maximal ideal of $\mathbf{T}_n$ generated by $\mathfrak{m}_\emptyset$ and $U_x - \alpha_x$ for $x \in Q$.

Let $\widetilde{\mathfrak{m}}_\alpha$ and $\widetilde{\mathfrak{m}}_\beta$ denote the ideals of $\widetilde{\mathbf{T}}_n$ containing $\mathfrak{m}$ and $U_p - \alpha$ or $U_p - \beta$ respectively. If $\alpha = \beta$, we simply write $\widetilde{\mathfrak{m}} = \widetilde{\mathfrak{m}}_\alpha = \widetilde{\mathfrak{m}}_\beta$. Note that since $n > 1$, we have

$$H^0\left(X_1(Q), \omega_{\mathcal{O}}^n\right) \otimes \mathcal{O}/\varpi^m \xrightarrow{\sim} H^0\left(X_1(Q), \omega_{\mathcal{O}/\varpi^m}^n\right)$$

and hence we may regard the latter as a module for $\widetilde{\mathbf{T}}_n$ (and its sub-algebras $\mathbf{T}_n$ and $\mathbf{T}_n^{\mathrm{an}}$). The proof of Theorem 3.11 will be completed in Sections 3.4–3.7. $\square$

3.4 Interlude: Galois representations in higher weight

In this section, we summarize some results about Galois representations associated to ordinary Hecke algebras in weight $n \geq 2$. As above, let α and β be the eigenvalues of $\overline{\rho}(\mathrm{Frob}_p)$. There is a natural map $\mathbf{T}_{n,\mathfrak{m}} \to \widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}_\alpha}$. If $\alpha = \beta$ this map is injective, otherwise, write $\mathbf{T}_{n,\mathfrak{m}_\alpha}$ for the image. There are continuous Galois representations

$$\rho_{n,\alpha} : G_{\mathbf{Q}} \to \mathrm{GL}_2\left(\mathbf{T}_{n,\mathfrak{m}_\alpha}\right)$$
$$\widetilde{\rho}_{n,\alpha} : G_{\mathbf{Q}} \to \mathrm{GL}_2\left(\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}_\alpha}\right)$$

with the following properties:

(a) The representation $\widetilde{\rho}_{n,\alpha}$ is obtained from $\rho_{n,\alpha}$ by composing $\rho_{n,\alpha}$ with the natural inclusion map $\mathbf{T}_{n,\mathfrak{m}_\alpha} \to \widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}_\alpha}$.

(b) $\rho_{n,\alpha}$ and $\widetilde{\rho}_{n,\alpha}$ are unramified at all primes $(x, pNQ) = 1$ and the characteristic polynomial of $\rho_{n,\alpha}(\mathrm{Frob}_x)$ for such x is

$$X^2 - T_x X + x^{n-1}\langle x \rangle.$$

(c) If E is a field of characteristic zero, and $\phi : \widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}} \to E$ is a homomorphism, then $\phi \circ \widetilde{\rho}_{n,\alpha} | G_p$ is equivalent to a representation of the form

$$\begin{pmatrix} \epsilon^{n-1}\lambda(\phi(\langle p \rangle)/\phi(U_p)) & * \\ 0 & \lambda(\phi(U_p)) \end{pmatrix}$$

where $\lambda(z)$ denotes the unramified character sending Frob_p to z.

These results follow from standard facts about Galois representations attached to classical ordinary modular forms, together with the fact that there is an inclusion

$$\mathbf{T}_{n,\mathfrak{m}_\alpha} \hookrightarrow \widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}_\alpha} \hookrightarrow \prod E_i$$

with E_i running over a finite collection of finite extensions of K corresponding to the ordinary eigenforms of weight n and level $\Gamma_1(NQ)$.

Theorem 3.14 *Under Assumption 3.9, $\rho_{n,\alpha}$ is a deformation of $\overline{\rho}$ that satisfies conditions (2) and (4) of Definition 3.1, with the exception of the condition of being unramified at p.*

This follows from the choice of N and H together with the local Langlands correspondence and results of Diamond–Taylor and Carayol (see [15, Lemma 5.1.1]). (The choice of H ensures that for each prime $x \neq p$, $\det(\rho_{n,\alpha})|I_x$ has order prime to p). Note that without Assumption 3.9, the representation $\rho_{n,\alpha}$ still satisfies condition (2) of Definition 3.1; the issue is that $\rho_{n,\alpha}$ may have extra ramification at those primes not in $T(\overline{\rho})$.

We now fix one of the eigenvalues of $\overline{\rho}(\mathrm{Frob}_p)$, α say, and write $\widetilde{\mathfrak{m}} = \widetilde{\mathfrak{m}}_\alpha$. The existence of $\widetilde{\rho}_{n,\alpha}$ gives $B := \widetilde{\mathbf{T}}^2_{n,\widetilde{\mathfrak{m}}}$ the structure of a $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}[G_{\mathbf{Q}}]$-module. Recall that G_p is the decomposition group of $G_{\mathbf{Q}}$ at p.

Lemma 3.15 *Suppose that $\overline{\rho}(\mathrm{Frob}_p)$ is not a scalar. Then there exists an exact sequence of $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}[G_p]$-modules*

$$0 \to A \to B \to C \to 0$$

such that:

1. *A and C are free $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}$-modules of rank one.*
2. *The sequence splits $B \simeq A \oplus C$ as a sequence of $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}$-modules.*
3. *The action of G_p on C factors through $G_{\mathbf{F}_p} = G_p/I_p$, and Frobenius acts via the operator $U_p \in \widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}$.*
4. *The action of G_p on A is unramified and is via the character $\epsilon^{n-1}\lambda(\langle p\rangle U_p^{-1})$.*

Proof Let C denote the maximal $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}$-quotient on which Frob_p acts by U_p. The construction of C is given by taking a quotient, and thus its formulation is preserved under taking quotients of B. Since B is free of rank two, $B/\widetilde{\mathfrak{m}}$ has dimension 2. The action of U_p on $C/\widetilde{\mathfrak{m}}$ is, by definition, given by the scalar α. Yet $B/\widetilde{\mathfrak{m}}$ as a G_p-representation is given by $\overline{\rho}$, and $\overline{\rho}(\mathrm{Frob}_p)$ either has distinct eigenvalues α and β or is non-scalar by definition. Hence $\dim C/\widetilde{\mathfrak{m}} = 1$, and by Nakayama's lemma, C is cyclic. On the other hand, by well-known properties of the Galois actions arising from classical modular forms, we see that $C \otimes \mathbf{Q}$ has rank one, and thus C is free as a $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}$-module. It follows that $B \to C$ splits, and that A is also free of rank one. Considering once more the local Galois structure of representations arising from classical modular forms, it follows that G_p acts on $A \otimes \mathbf{Q}$ via $\epsilon^{n-1}\lambda(\langle p\rangle U_p^{-1})$. Since $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}$ is $\mathcal{O}$-flat, the action of G_p on A itself is given by the same formula, proving the lemma. □

When $\overline{\rho}(\mathrm{Frob}_p)$ is scalar, there does not exist such a decomposition. Instead, in Sect. 3.7, we shall study the local properties of $\rho_{n,\alpha}$ and $\widetilde{\rho}_{n,\alpha}$ using finer properties of local deformation rings.

3.5 Proof of Theorem 3.11, case 1: $\alpha \neq \beta$, $\overline{\rho}(\mathrm{Frob}_p)$ has distinct eigenvalues

We claim there is an isomorphism of $\mathbf{T}^{\mathrm{univ}}$-modules:

$$H^0(X_1(Q), \omega_{\mathcal{O}/\varpi^m})_{\mathfrak{m}} \simeq (\psi H^0(X_1(Q), \omega_{\mathcal{O}/\varpi^m})^2)_{\widetilde{\mathfrak{m}}_\alpha}$$

obtained by composing ϕ with the $\widetilde{\mathbf{T}}_n$-equivariant projection onto $(\psi H^0(X_1(Q), \omega_{\mathcal{O}/\varpi^m})^2)_{\widetilde{\mathfrak{m}}_\alpha}$. The argument is similar to the proof of Lemma 3.5. We know that U_p satisfies the equation $X^2 - T_p X + \langle p\rangle = 0$ on the image of ψ but we may not use Hensel's Lemma to deduce that there exist $\widetilde{\alpha}$ and $\widetilde{\beta}$ in $\mathbf{T}_{\mathfrak{m}}/I_m$ such that $(U_p - \widetilde{\alpha})(U_p - \widetilde{\beta}) = 0$ on the $\mathfrak{m}$-part of the image of ψ, since we do not know *a priori* that T_p lies in $\mathbf{T}_{\mathfrak{m}}$. Instead, we note the following. Since U_p acts invertibly on the image of ψ, we deduce from the equality $T_p = U_p^{-1}\langle p\rangle + U_p$ that $T_p - \alpha - \beta$ lies in $\widetilde{\mathfrak{m}}_\alpha$ and $\widetilde{\mathfrak{m}}_\beta$, and thus acts nilpotently on $H^0(X_1(Q), \omega_{\mathcal{O}/\varpi^m})_{\mathfrak{m}}$. It follows that $\mathbf{T}_{\mathfrak{m}}/I_m[T_p] \subset \mathrm{End}(H^0(X_1(Q), \omega_{\mathcal{O}/\varpi^m})_{\mathfrak{m}})$ is a local ring with maximal ideal $\widetilde{\mathfrak{m}}$ which acts on $H^0(X_1(Q), \omega_{\mathcal{O}/\varpi^m})_{\mathfrak{m}}$. The operator U_p *does* satisfy the quadratic relation $X^2 - T_p X + \langle p\rangle = 0$ over $\mathbf{T}_{\mathfrak{m}}/I_m[T_p]$, and hence by Hensel's Lemma there exists $\widetilde{\alpha}$ and $\widetilde{\beta}$ in $\mathbf{T}_{\mathfrak{m}}/I_m[T_p]$ such that $(U_p - \widetilde{\alpha})(U_p - \widetilde{\beta}) = 0$ on the $\mathfrak{m}$-part of the image of ψ. The argument then proceeds as in the proof of Lemma 3.5, noting (tautologically) that $H^0(X_1(Q), \omega_{\mathcal{O}/\varpi^m})_{\widetilde{\mathfrak{m}}} = H^0(X_1(Q), \omega_{\mathcal{O}/\varpi^m})_{\mathfrak{m}}$.

It follows from the result just established that $\mathbf{T}_{\mathfrak{m}}/I_m[T_p][U_p] \subset \mathrm{End}(\mathrm{Im}(\psi)_{\widetilde{\mathfrak{m}}_\alpha})$ is a quotient of $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}_\alpha}$, and $\mathbf{T}_{\mathfrak{m}}/I_m$ is the corresponding quotient of $\mathbf{T}_{n,\mathfrak{m}_\alpha}$. Note that the trace of any lift of Frobenius on the corresponding quotient of $\rho_{n,\alpha}$ is equal to $U_p + \langle p\rangle U_p^{-1}$, which is equal to T_p in $\mathrm{End}(\mathrm{Im}(\psi)_{\widetilde{\mathfrak{m}}_\alpha})$. (We use here, as below, that ϵ^{n-1} is trivial modulo ϖ^m). In particular, this implies that $T_p \in \mathbf{T}_{\mathfrak{m}}/I_m$. We now define $\rho_{Q,m}$ to be the composition of $\rho_{n,\alpha}$ with the surjection $\mathbf{T}_{n,\mathfrak{m}_\alpha} \twoheadrightarrow \mathbf{T}_{\mathfrak{m}}/I_m$, and $\widetilde{\rho}_{Q,m}$ to be $\widetilde{\rho}_{n,\alpha}$ on the corresponding quotient $\mathbf{T}_{\mathfrak{m}}/I_m[U_p]$ of $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}_\alpha}$. (Since $U_p = \widetilde{\alpha}$ in $\mathbf{T}_{\mathfrak{m}}/I_m[U_p]$, the corresponding quotients of $\mathbf{T}_{n,\mathfrak{m}_\alpha}$ and $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}_\alpha}$ are the same). The character ν defined by the formula $\nu := \langle \det \overline{\rho} \rangle \det(\rho_{Q,m})^{-1}$ is thus unramified outside Q and of p-power order. Since $p > 2$, ν admits a square root η (also unramified outside Q). We have established that $\rho'_{Q,m} := \rho_{Q,m} \otimes \eta$ satisfies all the conditions of Definition 3.1, except the condition that it be unramified at p. Equivalently, it suffices to show that $\widetilde{\rho}_{Q,m}$ is unramified at p. By Lemma 3.15, we may write

$$\widetilde{\rho}_{Q,m}|G_p \cong \begin{pmatrix} \lambda(\widetilde{\beta}) & * \\ 0 & \lambda(\widetilde{\alpha}) \end{pmatrix}.$$

By symmetry, we could equally well have defined $\rho_{Q,m}$ by regarding $\mathbf{T}_{\mathfrak{m}}/I_m$ as a quotient of $\mathbf{T}_{n,\mathfrak{m}_\beta}$. (Note that the Chebotarev density theorem and [39, Théorème 1] imply that $\rho_{Q,m}$ is uniquely determined by the condition that $\mathrm{Trace}\rho_{Q,m}(\mathrm{Frob}_x) = T_x$ for all $(x, pNQ) = 1$). It follows that we also have

$$\widetilde{\rho}_{Q,m}|G_p \cong \begin{pmatrix} \lambda(\widetilde{\alpha}) & * \\ 0 & \lambda(\widetilde{\beta}) \end{pmatrix}.$$

Since $\alpha \neq \beta$, this forces $\rho_{Q,m}|G_p$ to split as a direct sum of the unramified character $\lambda(\widetilde{\alpha})$ and $\lambda(\widetilde{\beta})$. (Moreover, we see that $T_p = U_p^{-1}\langle p \rangle + U_p = \widetilde{\alpha} + \widetilde{\beta} = \mathrm{Trace}(\rho_{Q,m}(\mathrm{Frob}_p)) \in \mathbf{T}_{\mathfrak{m}}$).

3.6 Proof of Theorem 3.11, case 2: $\alpha = \beta$, $\overline{\rho}(\mathrm{Frob}_p)$ non-scalar

We will assume below that $\alpha = \beta$ is a generalized eigenvalue of $\overline{\rho}(\mathrm{Frob}_p)$, and furthermore that $\overline{\rho}(\mathrm{Frob}_p)$ is non-scalar. However, we first prove the lemma below.

Lemma 3.16 *(Doubling) Without any assumption on* $\overline{\rho}(\mathrm{Frob}_p)$*, the action of* $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}$ *on* $\psi(H^0(X_1(Q), \omega_{\mathcal{O}/\varpi^m})^2_{\mathfrak{m}})$ *factors through a quotient isomorphic to*

$$\mathbf{T}_{\mathfrak{m}}/I_m[T_p][X]/(X^2 - T_p X + \langle p \rangle),$$

where U_p *acts by* X.

Proof The action of $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}$ certainly contains $T_p := U_p + \langle p \rangle U_p^{-1}$, and moreover U_p also satisfies the indicated relation. Thus it suffices to show that U_p does not satisfy any *further* relation. Such a relation would be of the form $AU_p + B = 0$ for operators A, B in $\mathbf{T}_{\mathfrak{m}}/I_m[T_p]$. By considering the action of U_p as a matrix on the image of ψ, however, this would imply an identity:

$$A \begin{pmatrix} T_p & 1 \\ -\langle p \rangle & 0 \end{pmatrix} + B \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

from which one deduces that $A = B = 0$ (the fact that one can deduce the vanishing of the entries is equivalent to the injectivity of ψ). □

Following Wiese [38], we call this phenomenon "doubling", because the corresponding quotient of $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}$ contains two copies of the image of $\mathbf{T}_{n,\mathfrak{m}}$. We show that this implies that the corresponding Galois representation is unramified.

Note that the trace under $\rho_{n,\alpha}$ of any lift of Frobenius is sent to $U_p + \langle p \rangle U_p^{-1} = T_p$ in $\mathbf{T}_{\mathfrak{m}}/I_m$, and so this in particular implies that $T_p \in \mathbf{T}_{\mathfrak{m}}/I_m$. The image of $\mathbf{T}_{n,\mathfrak{m}}$ under the map

$$\mathbf{T}_{n,\mathfrak{m}} \to \widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}} \twoheadrightarrow \mathbf{T}_{\mathfrak{m}}/I_m[T_p][U_p]$$

is given by $\mathbf{T}_{\mathfrak{m}}/I_m = \mathbf{T}_{\mathfrak{m}}/I_m[T_p]$. We thus obtain a Galois representation

$$\rho_{Q,m} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{T}_{\mathfrak{m}}/I_m).$$

As in Sect. 3.5, it suffices to prove that $\rho_{Q,m}$ is unramified at p. Consider the Galois representation $\widetilde{\rho}_{Q,m} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{T}_{\mathfrak{m}}/I_m[U_p])$ obtained by tensoring over $\mathbf{T}_{n,\mathfrak{m}}$ with $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}$. By Lemma 3.16, there is an isomorphism

$$\mathbf{T}_{\mathfrak{m}}/I_m[U_p] \simeq \mathbf{T}_{\mathfrak{m}}/I_m \oplus \mathbf{T}_{\mathfrak{m}}/I_m$$

as a $\mathbf{T}_{\mathfrak{m}}/I_m$-module. Since $\widetilde{\rho}_{Q,m}$ is obtained from $\rho_{Q,m}$ by tensoring with a doubled module, it follows that there is an isomorphism $\widetilde{\rho}_{Q,m} \simeq \rho_{Q,m} \oplus \rho_{Q,m}$ as a $\mathbf{T}_{\mathfrak{m}}/I_m[G_p]$-module (or even $\mathbf{T}_{\mathfrak{m}}/I_m[G_{\mathbf{Q}}]$-module).

Lemma 3.17 *Let $(R, \mathfrak{m})$ be a local ring, and let N, M, and L be $R[G_p]$-modules which are free R-modules of rank two. Suppose there is an exact sequence of $R[G_p]$-modules*

$$0 \to N \to M \oplus M \to L \to 0$$

which is split as a sequence of R-modules. Suppose that $L/\mathfrak{m}$ is indecomposable as a $R[G_p]$-module. Then $N \simeq M \simeq L$ as $R[G_p]$-modules.

Proof This is Proposition 4.4 of Wiese [38] (we use L here instead of Q in [38] to avoid notational conflicts). Note that the lemma is stated for $\mathbf{F}_p$-algebras R and the stated condition is on the sub-module N rather than the quotient L, but the proof is exactly the same. □

We apply this as follows. Consider the sequence of $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}$-modules considered in Lemma 3.15. If we tensor this sequence with the quotient of $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}$ corresponding to the doubling isomorphism

$$\mathbf{T}_{\mathfrak{m}}/I_m[U_p] = \mathbf{T}_{\mathfrak{m}}/I_m \oplus \mathbf{T}_{\mathfrak{m}}/I_m,$$

then the corresponding quotient of B is $\widetilde{\rho}_{Q,m}$, which, by doubling, is free of rank 4 over $\mathbf{T}_{\mathfrak{m}}/I_m$ and as a $\mathbf{T}_{\mathfrak{m}}/I_m[G_p]$-module is given by $\rho_{Q,m} \oplus \rho_{Q,m}$. The corresponding quotients A and C are similarly free over $\mathbf{T}_{\mathfrak{m}}/I_m$ of rank 2. The action of Frob_p on $L/\mathfrak{m}$ is given by U_p. Since U_p does not lie in $\mathfrak{m}$—as this would contradict doubling—it follows that $(U_p - \alpha)$ acts nilpotently but non-trivially on $L/\mathfrak{m}$, and hence $L/\mathfrak{m}$ is indecomposable (indeed, by construction, $L/\mathfrak{m}$ is free of rank one over $k[U_p]/(U_p - \alpha)^2$). Hence, by the lemma above, there are isomorphisms $L \simeq \rho_{Q,m}$ as a G_p-module. Yet L is a quotient of C, which is by construction unramified, and thus $\rho_{Q,m}$ is also unramified. Finally, we note that the trace of Frobenius at p is given by $U_p + \langle p \rangle U_p^{-1} = T_p$, so $T_p = \mathrm{Trace}(\rho_{Q,m}(\mathrm{Frob}_p))$.

3.7 Proof of Theorem 3.11, case 3: $\boldsymbol{\alpha = \beta}$, $\boldsymbol{\overline{\rho}}$(Frob$_{\boldsymbol{p}}$) scalar

The construction of the previous section gives a representation $\rho_{Q,m}$ which satisfies all the required deformation properties with the possible exception of knowing that $\rho_{Q,m}$ is unramified at p. In order to deal with the case when $\overline{\rho}(\mathrm{Frob}_p)$ is scalar, we shall have to undergo a closer study of local deformation rings. Suppose that

$$\overline{\rho} : G_p \to \mathrm{GL}_2(k)$$

is trivial. (If $\overline{\rho}$ is scalar, it is trivial after twisting). We introduce some local framed universal deformation rings associated to $\overline{\rho}$. Fix a lift $\phi_p \in G_p$ of Frob_p.

In the definition below, an *eigenvalue* of a linear operator is defined to be a root of the corresponding characteristic polynomial.

Definition 3.18 For A in $\mathcal{C}_{\mathcal{O}}$, let $D(A)$ denote the set of framed deformations of $\overline{\rho}$ to A, and let $\widetilde{D}(A)$ denote the framed deformations together with an eigenvalue α of ϕ_p. Let these functors be represented by rings R^{univ} and $\widetilde{R}^{\mathrm{univ}}$ respectively.

There is a natural inclusion $R^{\mathrm{univ}} \to \widetilde{R}^{\mathrm{univ}}$ and $\widetilde{R}^{\mathrm{univ}}$ is isomorphic to a quadratic extension of R^{univ} (given by the characteristic polynomial of ϕ_p). Kisin constructs certain quotients of R^{univ} which capture characteristic zero quotients with good p-adic Hodge theoretic properties. Let ϵ denote the cyclotomic character, let ω denote the Teichmüller lift of the mod-p reduction of ϵ, and let $\chi = \epsilon\omega^{-1}$, so $\chi \equiv 1 \mod \varpi$. We modify the choice of ϕ_p if necessary so that $\chi(\phi_p) = 1$. Let $R^{\mathrm{univ},\chi^{n-1}}$ and $\widetilde{R}^{\mathrm{univ},\chi^{n-1}}$ denote the quotients of R^{univ} and $\widetilde{R}^{\mathrm{univ}}$ corresponding to deformations with determinant χ^{n-1}.

Theorem 3.19 *Fix an integer $n \geq 2$.*

(1) *There exists a unique reduced $\mathcal{O}$-flat quotient $\widetilde{R}^{\dagger}$ of $\widetilde{R}^{\mathrm{univ},\chi^{n-1}}$ such that points on the generic fiber of $\widetilde{R}^{\dagger}$ correspond to representations $\rho : G_p \to \mathrm{GL}_2(E)$ such that:*

$$\rho \sim \begin{pmatrix} \chi^{n-1}\lambda\left(\alpha^{-1}\right) & * \\ 0 & \lambda(\alpha) \end{pmatrix}$$

(2) *The ring $\widetilde{R}^{\dagger}$ is an integral domain which is normal, Cohen–Macaulay, and of relative dimension 4 over $\mathcal{O}$.*

Proof Part 1 of this theorem is due to Kisin. For part 2, the fact that $\widetilde{R}^{\dagger}$ is an integral domain follows from the proof of Lemma 3.4.3 of [40]. The rest of part 2 follows from the method and results of Snowden [41]. More precisely, Snowden works over an arbitrary finite extension of $\mathbf{Q}_p$ containing $\mathbf{Q}_p(\zeta_p)$, and assumes that $n = 2$, so $\chi = \epsilon\omega^{-1} = \epsilon$. However, this is exactly the hardest case—since for us $p \neq 2$, $\chi^{n-1} \neq \epsilon$, our deformation problem consists of a single potentially crystalline component. In particular, the arguments of [41] show that $\widetilde{R}^{\dagger} \otimes k$ is an integral normal Cohen–Macaulay ring of dimension four, which is not Gorenstein, and is identified (in the notation of *ibid*). with the completion of $\mathcal{B}_1$ at $b = (1, 1; 0)$. □

Let $R^{\dagger}$ denote the image of R^{univ} in $\widetilde{R}^{\dagger}$. We also define the following rings:

Definition 3.20 Let R^{unr} denote the largest quotient of $R^{\dagger}$ corresponding to unramified deformations of $\overline{\rho}$. Let $\widetilde{R}^{\mathrm{unr}}$ denote the corresponding quotient of $\widetilde{R}^{\dagger}$.

We are now in a position to define two ideals of R^{univ}.

Definition 3.21 The *unramified ideal* $\mathscr{I}$ is the kernel of the map $R^{\mathrm{univ}} \to R^{\mathrm{unr}}$. The *doubling ideal* $\mathscr{J}$ is the annihilator of $\widetilde{R}^{\dagger}/R^{\dagger}$ as an R^{univ}-module.

Lemma 3.22 *There is an equality $\mathscr{J} = \mathscr{I}$.*

Proof We first prove the inclusion $\mathscr{J} \subseteq \mathscr{I}$. By definition, $R^{\mathrm{univ}}/\mathscr{J}$ acts faithfully on $\widetilde{R}^{\dagger}/R^{\dagger}$, and it is the largest such quotient. Hence it suffices to show that $R^{\mathrm{univ}}/\mathscr{I}$ acts faithfully on

$$(\widetilde{R}^{\dagger}/R^{\dagger}) \otimes R^{\mathrm{univ}}/\mathscr{I} \simeq (\widetilde{R}^{\dagger}/\mathscr{I})/(R^{\dagger}/\mathscr{I}).$$

Since $\widetilde{R}^{\dagger}/\mathscr{I} \simeq \widetilde{R}^{\mathrm{unr}}$ and $R^{\dagger}/\mathscr{I} \simeq R^{\mathrm{unr}}$, it suffices to show that $\widetilde{R}^{\mathrm{unr}}/R^{\mathrm{unr}}$ is a faithful $R^{\mathrm{unr}} = R^{\mathrm{univ}}/\mathscr{I}$-module. (It is not *a priori* obvious that the map $R^{\mathrm{unr}} \to \widetilde{R}^{\mathrm{unr}}$ is injective, so the notation $\widetilde{R}^{\mathrm{unr}}/R^{\mathrm{unr}}$ is slightly misleading;

however, we prove it is so by explicit computation below). Let ϖ^m denote the greatest power of ϖ dividing $(\chi^{n-1}(g) - 1)\mathcal{O}$ for all g in the decomposition group at p. By considering determinants, R^{unr} (and $\widetilde{R}^{\mathrm{unr}}$) is annihilated by ϖ^m. The moduli space of matrices ϕ in $\mathcal{O}/\varpi^m$ which are trivial modulo ϖ and have determinant one, that is, with

$$\phi = \begin{pmatrix} 1+\phi_1 & \phi_2 \\ \phi_3 & 1+\phi_4 \end{pmatrix},$$

is represented by:

$$\mathcal{O}/\varpi^m[\phi_1, \phi_2, \phi_3, \phi_4]/(\phi_1 + \phi_4 + \phi_1\phi_4 - \phi_2\phi_3).$$

We show shortly that this ring is isomorphic to R^{unr}; admit this for a moment. The corresponding moduli space $\widetilde{R}^{\mathrm{unr}}$ of such matrices together with an eigenvalue $\alpha = 1 + \beta$ is represented by

$$\widetilde{R}^{\mathrm{unr}} \simeq R^{\mathrm{unr}}[\beta]/(\beta^2 - (\phi_1 + \phi_4)\beta - (\phi_1 + \phi_4)) \simeq R^{\mathrm{unr}} \oplus R^{\mathrm{unr}},$$

where the last isomorphism is as an R^{unr}-module. Clearly R^{unr} acts faithfully on $(R^{\mathrm{unr}} \oplus R^{\mathrm{unr}})/R^{\mathrm{unr}} \simeq R^{\mathrm{unr}}$, proving the inclusion $\mathscr{J} \subseteq \mathscr{I}$. We now prove the equality of rings above. It suffices to prove it for $\widetilde{R}^{\mathrm{unr}}$. By construction, the ring above certainly surjects onto $\widetilde{R}^{\mathrm{unr}}$. Hence, it suffices to show that this ring is naturally a quotient of $\widetilde{R}^{\dagger}$. As in Snowden, the ring $\widetilde{R}^{\dagger}$ represents the functor given by deformations to A with eigenvalue α satisfying the following equations:

(1) $\phi \in M_2(A)$ has determinant 1.
(2) α is a root of the characteristic polynomial of ϕ.
(3) $\mathrm{Trace}(g) = \chi^{n-1}(g) + 1$ for $g \in I_p$.
(4) $(g-1)(g'-1) = (\chi^{n-1}(g) - 1)(g'-1)$ for $g, g' \in I_p$.
(5) $(g-1)(\phi - \alpha) = (\chi^{n-1}(g) - 1)(\phi - \alpha)$ for $g \in I_p$.
(6) $(\phi - \alpha)(g-1) = (\alpha^{-1} - \alpha)(g-1)$ for $g \in I_p$.

To understand where these equations come from, one should imagine writing down the following equations:

$$\rho(\phi_p) = \phi \approx \begin{pmatrix} \alpha^{-1} & * \\ 0 & \alpha \end{pmatrix}, \qquad \rho(g) \approx \begin{pmatrix} \chi^{n-1}(g) & * \\ 0 & 1 \end{pmatrix}, \ g \in I_p.$$

We caution, however, that although one can find such a basis for any representation when A is a field, we do not claim that there exists any universal such basis (indeed, we presume that there does not). Returning to our argument, it is now trivial to observe that the quotient of $\widetilde{R}^{\dagger}$ in which $g \in I_p$ is the identity is equal to the ring we asserted to be $\widetilde{R}^{\mathrm{unr}}$ above (and that the image of R^{univ} in this ring is what we asserted to be R^{unr}).

We now prove the opposite inclusion, namely that $\mathscr{I} \subseteq \mathscr{J}$. Instead of writing down a presentation of $\widetilde{R}^{\dagger}$, it will suffice to note the following, which follows from the explicit description above: The ring $\widetilde{R}^{\dagger}$ is generated over $\mathcal{O}$ by the following elements which all lie in the maximal ideal:

(1) Parameters ϕ_i (for $i = 1$ to 4) corresponding to the image of $\rho(\phi_p) - 1$,
(2) Parameters x_{ij} for $i = 1$ to 4 and a finite number of j corresponding to the image of inertial elements $m_j = \rho(g_j) - 1$.
(3) An element β, where $\alpha = 1 + \beta$ is an eigenvalue of $\rho(\phi_p)$.

Moreover, $R^{\dagger}$ is generated as a sub-algebra by ϕ_i and the x_{ij}, and β satisfies

$$\beta^2 - (\phi_1 + \phi_4)\beta - (\phi_1 + \phi_4) = 0.$$

Since the determinant of ϕ is one, it follows that $\alpha + \alpha^{-1} = 2 + \phi_1 + \phi_4$. By definition, there is a decomposition of $R^{\dagger}$-modules $\widetilde{R}^{\dagger}/\mathscr{J} = R^{\dagger}/\mathscr{J} \oplus \beta R^{\dagger}/\mathscr{J}$ with each summand being free over $R^{\dagger}/\mathscr{J}$. From the equality (6), we deduce that the relation

$$\begin{aligned}(\phi - 1)m_j - (\phi_1 + \phi_4)m_j = (\alpha^{-1} - 1 - \phi_1 - \phi_4)m_j &= -(\alpha - 1)m_j \\ &= -\beta m_j\end{aligned}$$

holds in $M_2(\widetilde{R}^{\dagger})$, and hence also in $M_2(\widetilde{R}^{\dagger}/\mathscr{J})$. Yet by assumption, over $R^{\dagger}/\mathscr{J}$, the modules $R^{\dagger}/\mathscr{J}$ and $\beta R^{\dagger}/\mathscr{J}$ have trivial intersection, from which it follows that $\beta m_j = 0$ in $M_2(\beta R^{\dagger}/\mathscr{J})$. In particular, since the latter module is generated by β, we must have $x_{ij} \in \mathscr{J}$ for all i and j. Since $\mathscr{I}$ is generated by x_{ij}, we deduce that $\mathscr{I} \subset \mathscr{J}$, and hence that $\mathscr{I} = \mathscr{J}$. □

Remark 3.23 Why might one expect an equality $\mathscr{I} = \mathscr{J}$? One reason is as follows. The doubling ideal $\mathscr{J}$ represents the largest quotient of $\widetilde{R}^{\dagger}$ on which the eigenvalue of Frobenius α *cannot be distinguished from its inverse* α^{-1}. Slightly more precisely, it is the largest quotient for which there is an isomorphism $\widetilde{R}^{\dagger}/\mathscr{J} \to \widetilde{R}^{\dagger}/\mathscr{J}$ fixing the image of $R^{\dagger}$ and sending α to α^{-1}. It is clear that such an isomorphism exists for unramified representations. Similarly, for a ramified ordinary quotient, one might expect that the α can be distinguished from α^{-1} by looking at the "unramified quotient line" of the representation. Indeed, for characteristic zero representations this is clear—one even has $R^{\dagger}[1/\varpi] \simeq \widetilde{R}^{\dagger}[1/\varpi]$.

Lemma 3.24 *There is a surjection* $\mathbf{T}_{n,\mathfrak{m}} \otimes_{R^{\dagger}} \widetilde{R}^{\dagger} \to \widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}$.

Proof Recall that $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}} = \mathbf{T}_{n,\mathfrak{m}}[U_p]$. Since U_p is given as an eigenvalue of Frobenius, $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}$ naturally has the structure of a $\widetilde{R}^{\mathrm{univ}}$-algebra. We claim that the map from $\widetilde{R}^{\mathrm{univ}}$ to $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}$ factors through $\widetilde{R}^{\dagger}$. Since $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}$ acts faithfully on a space of modular forms, there is an injection:

$$\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}} \hookrightarrow \prod E_i$$

into a product of fields corresponding to the Galois representations associated to the ordinary modular forms of weight n and level $\Gamma_1(NQ)$. By the construction of $\widetilde{R}^{\dagger}$, it follows that the map from $\widetilde{R}^{\mathrm{univ}}$ to this product factors through via $\widetilde{R}^{\dagger}$. This also implies that the map from R^{univ} to $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}$ (and hence to $\mathbf{T}_{n,\mathfrak{m}}$) factors through $R^{\dagger}$, and hence there exists a map

$$\mathbf{T}_{n,\mathfrak{m}} \otimes_{R^{\dagger}} \widetilde{R}^{\dagger} \rightarrow \widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}},$$

sending $\alpha \in \widetilde{R}^{\dagger}$ to U_p. Yet the image of this map contains both $\mathbf{T}_{n,\mathfrak{m}}$ and U_p, and is thus surjective. □

Definition 3.25 Let the *global doubling ideal* $\mathscr{J}^{\mathrm{glob}}$ be the annihilator of $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}/\mathbf{T}_{n,\mathfrak{m}}$ as an $R^{\dagger}$-module.

Since there is a surjection $\mathbf{T}_{n,\mathfrak{m}} \otimes_{R^{\dagger}} \widetilde{R}^{\dagger} \rightarrow \widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}$, it follows that $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}/\mathbf{T}_{n,\mathfrak{m}}$ is a quotient of

$$(\mathbf{T}_{n,\mathfrak{m}} \otimes_{R^{\dagger}} \widetilde{R}^{\dagger})/\mathbf{T}_{n,m} = (\mathbf{T}_{n,\mathfrak{m}} \otimes_{R^{\dagger}} \widetilde{R}^{\dagger})/\mathbf{T}_{n,m} \otimes_{R^{\dagger}} R^{\dagger} \simeq \mathbf{T}_{n,\mathfrak{m}} \otimes_{R^{\dagger}} \widetilde{R}^{\dagger}/R^{\dagger}$$

as an $R^{\dagger}$-module. In particular, by considering the action on the last factor, we deduce that $\mathscr{J} \subset \mathscr{J}^{\mathrm{glob}}$. In particular, $\mathscr{I} \subset \mathscr{J}^{\mathrm{glob}}$, or equivalently, on any quotient of $\mathbf{T}_{n,\mathfrak{m}}$ on which the corresponding quotient of $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}$ is doubled (in the sense that the quotient of $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}$ is free of rank 2 as a module for the image of $\mathbf{T}_{n,\mathfrak{m}}$), the action of the Galois group at p is unramified. In particular, by Lemma 3.16, this applies to the quotient of $\widetilde{\mathbf{T}}_{n,\widetilde{\mathfrak{m}}}$ given by $\mathbf{T}_{\mathfrak{m}}/I_m[T_p][U_p]$. Specifically, as in the previous sections, we obtain corresponding Galois representations:

$$\rho_{Q,m} : G_{\mathbf{Q}} \rightarrow \mathrm{GL}_2(\mathbf{T}_{\mathfrak{m}}/I_m), \qquad \widetilde{\rho}_{Q,m} : G_{\mathbf{Q}} \rightarrow \mathrm{GL}_2(\mathbf{T}_{\mathfrak{m}}/I_m[U_p]).$$

(The trace of any lift of Frobenius on this quotient is equal to $U_p + \langle p \rangle U_p^{-1} = T_p$, and so $T_p \in \mathbf{T}_{\mathfrak{m}}/I_m$). From the discussion above, we deduce that $\widetilde{\rho}_{Q,m}$ and thus $\rho_{Q,m}$ is unramified at p, and that $\mathrm{Trace}(\rho_{Q,m}(\mathrm{Frob}_p)) = T_p$. The rest of the argument follows as in Case 3.5, and this completes the proof of Theorem 3.11 □

3.8 Modularity lifting

We now return to the situation of Sect. 3.1. Taking $Q = 1$ in Theorem 3.11, we obtain a minimal deformation $\rho' : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{T}_{\emptyset,\mathfrak{m}_\emptyset})$ of $\overline{\rho}$ and hence a homomorphism $\varphi : R^{\min} \to \mathbf{T}_{\emptyset,\mathfrak{m}_\emptyset}$ which is easily seen to be surjective. Recall that Assumption 3.9 is still in force.

Theorem 3.26 *The map $\varphi : R^{\min} \twoheadrightarrow \mathbf{T}_{\emptyset,\mathfrak{m}_\emptyset}$ is an isomorphism and $\mathbf{T}_{\emptyset,\mathfrak{m}_\emptyset}$ acts freely on $H_0(X,\omega)_{\mathfrak{m}_\emptyset}$.*

Proof We view $H_0(X,\omega)_{\mathfrak{m}_\emptyset}$ as an $R^{\min}$-module via φ. Since φ is surjective, to prove the theorem, it suffices to show that $H_0(X,\omega)_{\mathfrak{m}_\emptyset}$ is free over $R^{\min}$. To show this, we will apply Proposition 2.3.

We set $R = R^{\min}$ and $H = H_0(X,\omega)_{\mathfrak{m}_\emptyset}$ and we define

$$q := \dim_k H^1_\emptyset\left(G_{\mathbf{Q}}, \mathrm{ad}^0\overline{\rho}(1)\right).$$

Note that $q \geq 1$ by Proposition 3.3. As in Proposition 2.3, we set $S_N = \mathcal{O}[(\mathbf{Z}/p^N\mathbf{Z})^q]$ for each integer $N \geq 1$ and we let R_∞ denote the power series ring $\mathcal{O}[x_1,\ldots,x_{q-1}]$. For each integer $N \geq 1$, fix a set of primes Q_N of $\mathbf{Q}$ satisfying the properties of Proposition 3.3. We can and do fix a surjection $\widetilde{\phi}_N : R_\infty \twoheadrightarrow R_{Q_N}$ for each $N \geq 1$. We let ϕ_N denote the composition of $\widetilde{\phi}_N$ with the natural surjection $R_{Q_N} \twoheadrightarrow R^{\min}$. Let

$$\Delta_{Q_N} = \prod_{x \in Q_N} (\mathbf{Z}/x)^\times$$

and choose a surjection $\Delta_{Q_N} \twoheadrightarrow \Delta_N := (\mathbf{Z}/p^N\mathbf{Z})^q$. Let $X_{\Delta_N}(Q_N) \to X_0(Q_N)$ denote the corresponding Galois cover. For each $x \in Q_N$, choose an eigenvalue α_x of $\overline{\rho}(\mathrm{Frob}_x)$. We let $\mathbf{T}_{Q_N}$ denote the Hecke algebra denoted $\mathbf{T}$ in Sect. 3.2.4 with the Q of that section taken to be the current Q_N. We let $\mathfrak{m}$ denote the maximal ideal of $\mathbf{T}_{Q_N}$ generated by $\mathfrak{m}_\emptyset$ and $U_x - \alpha_x$ for each $x \in Q_N$. We set $H_N := H_0(X_{\Delta_N}(Q_N), \omega_{\mathcal{O}})_{\mathfrak{m}}$. Then H_N is naturally an $\mathcal{O}[\Delta_N] = S_N$-module. By Theorem 3.11, we deduce the existence of a surjective homomorphism $R_{Q_N} \twoheadrightarrow \mathbf{T}_{Q_N,\mathfrak{m}}$. Since $\mathbf{T}_{Q_N,\mathfrak{m}}$ acts on H_N, we get an induced action of R_∞ on H_N (via $\widetilde{\phi}_N$ and the map $R_{Q_N} \twoheadrightarrow \mathbf{T}_{Q_N,\mathfrak{m}}$). We can therefore view H_N as a module over $R_\infty \otimes_{\mathcal{O}} S_N$.

To apply Proposition 2.3, it remains to check points (5a)–(5c). We check these conditions one by one:

(a) The image of S_N in $\mathrm{End}_{\mathcal{O}}(H_N)$ is contained in the image of R_∞ by construction (see Theorem 3.11). The second part of condition 5a is a consequence of the following: for each $x \in Q_N$, the restriction to G_x of the universal representation $G_{\mathbf{Q}} \to \mathrm{GL}_2(R_{Q_N})$ is of the form $\chi_{\alpha_x} \oplus \chi_{\beta_x}$ where each summand is of rank 1 over R_{Q_N} and where χ_{α_x} lifts $\lambda(\alpha_x)$. By restricting χ_{α_x} to I_x for each $x \in Q_N$, we obtain, by local class field theory, a map $\mathcal{O}[\Delta_{Q_N}] \to R_{Q_N}$. The quotient of R_{Q_N} by the image of the augmentation ideal of $\mathcal{O}[\Delta_{Q_N}]$ is just $R^{\min}$.

(b) As in the proof of Proposition 3.8, we have a Hochschild–Serre spectral sequence

$$\mathrm{Tor}_i^{S_N}(H_j(X_{\Delta_N}(Q_N), \omega)_{\mathfrak{m}}, \mathcal{O}) \implies H_{i+j}(X_0(Q_N), \omega)_{\mathfrak{m}}.$$

We see that $(H_N)_{\Delta_N} \cong H_0(X_0(Q_N), \omega)_{\mathfrak{m}}$. Then, by Lemmas 3.5 and 3.7 we obtain an isomorphism $(H_N)_{\Delta_N} \cong H_0(X, \omega)_{\mathfrak{m}_\emptyset} = H$, as required.

(c) The module H_N is finite over $\mathcal{O}$ and hence over S_N. Proposition 3.8 implies that $d_{S_N}(H_N) \geq 0$.

We may therefore apply Proposition 2.3 to deduce that H is free over R, completing the proof. □

We now deduce Theorem 1.4, under Assumption 3.9, from the previous result. In the statement of Theorem 1.4, we take $X_U = X = X_1(N)/H$ and $\mathcal{L}_\sigma = \mathcal{O}_X$ and, as in the statement, we let $\mathbf{T}$ be the Hecke algebra of $H^1(X, \omega)$ (generated by prime-to-Np Hecke operators) and $\mathfrak{m}$ the maximal ideal of $\mathbf{T}$ corresponding to $\overline{\rho}$. Analogous to the discussion preceding Proposition 3.8), we have a Hecke equivariant isomorphism $H_0(X, \omega) \xrightarrow{\sim} H^1(X, \omega)$ which thus gives rise to an isomorphism $\mathbf{T}_{\emptyset, \mathfrak{m}_\emptyset} \xrightarrow{\sim} \mathbf{T}_{\mathfrak{m}}$.

We also show that $H_0(X, \omega)_{\mathfrak{m}_\emptyset}$ has rank one as a $\mathbf{T}_{\emptyset, \mathfrak{m}_\emptyset}$-module: this follows by multiplicity one for $\mathrm{GL}(2)/\mathbf{Q}$ if $H_0(X, \omega_K)_{\mathfrak{m}_\emptyset}$ is non-zero. In the finite case, we argue as follows. By Nakayama's lemma it suffices to show that $H^0(X, \omega_k(-\infty))[\mathfrak{m}_\emptyset]$ has dimension one. We claim that $U_x \in \mathbf{T}_{\emptyset, \mathfrak{m}_\emptyset}$ for all $x|N$. This is a consequence of the assumption that $N(\overline{\rho}) = N$ as we now explain. Suppose that $x \| N$. Then the $\mathbf{T}^2_{\mathfrak{m}}$-representation has a unique invariant $\mathbf{T}_{\mathfrak{m}}$-line on which Frob_x acts by U_x, and so $U_x \in \mathbf{T}_{\mathfrak{m}}$. On the other hand, if $x^2|N$, then $U_x = 0$ is also in $\mathbf{T}_{\mathfrak{m}}$. Since we have also shown that $T_p \in \mathbf{T}_{\emptyset, \mathfrak{m}_\emptyset}$, we may deduce this from the fact that q-expansion is completely determined by the Hecke eigenvalues T_x for all $(x, N) = 1$ and U_x for all $x|N$.

3.9 Vexing primes

In this section, we detail the modifications to the previous arguments which are required to deal with vexing primes. To recall the difficulty, recall that a prime x different from p is **vexing** if:

(1) $\overline{\rho}|D_x$ is absolutely irreducible.
(2) $\overline{\rho}|I_x \simeq \xi \oplus \xi^c$ is reducible.
(3) $x \equiv -1 \mod p$.

The vexing nature of these primes can be described as follows: in order to realize $\overline{\rho}$ automorphically, one must work with $\Gamma_1(x^n)$ structure where x^n is the Artin conductor of $\overline{\rho}|D_x$. However, according to local Langlands, at such a level we also expect to see *non*-minimal deformations of $\overline{\rho}$, namely, deformations with $\rho|I_x \simeq \psi\langle\xi\rangle \oplus \psi^{-1}\langle\xi^c\rangle$, where ψ is a character of $(\mathbf{F}_{x^2})^\times$ of p-power order. Diamond [13] was the first to address this problem by observing that one can cut out a smaller space of modular forms by using the local Langlands correspondence. The version of this argument in [15] can be explained as follows. By Shapiro's Lemma, working with trivial coefficients at level $\Gamma(x^n)$ is the same as working at level prime to x where one now replaces trivial coefficients $\mathbf{Z}$ by a local system $\mathcal{F}$ corresponding to the group ring of the corresponding geometric cover. In order to avoid non-minimal lifts of $\overline{\rho}$, one works with a smaller local system $\mathcal{F}_\sigma$ cut out of $\mathcal{F}$ by a representation σ of the Galois group of the cover to capture exactly the minimal automorphic lifts of $\overline{\rho}$. The representation σ corresponds to a fixed inertial type at x. In our setting (coherent cohomology) we may carry out a completely analogous construction. Thus, instead, we shall construct a vector bundle $\mathcal{L}_\sigma$ on X. We then replace $H^*(X_1(N), \omega)$ by the groups $H^*(X_1(N), \omega \otimes \mathcal{L}_\sigma)$. The main points to check are as follows:

(1) The spaces $H^0(X_1(N), \omega^{\otimes n} \otimes \mathcal{L}_\sigma)$ for $n \geq 1$ do indeed cut out the requisite spaces of automorphic forms.
(2) This construction is sufficiently functorial so that all the associated cohomology groups admit actions by Hecke operators.
(3) These cohomology groups inject into natural spaces of q-expansions.
(4) This construction is compatible with arguments involving the Hochschild–Serre spectral sequence and Verdier duality.

We start by discussing some more refined properties of modular curves, in the spirit of Sect. 3.2.1. Let $S(\overline{\rho})$, $T(\overline{\rho})$ and Q be as in Sect. 3.1. Let $P(\overline{\rho})$ denote the set of $x \in S(\overline{\rho}) - T(\overline{\rho})$ where $\overline{\rho}$ is ramified and reducible.

We will now introduce compact open subgroups $V \lhd U \subset \mathrm{GL}_2(\mathbf{A}^\infty)$ and later we will fix a representation σ of U/V on a finite free $\mathcal{O}$ module W_σ. (In applications, U, V and σ will be chosen to capture all minimal modular lifts of $\overline{\rho}$. If the set of vexing primes $T(\overline{\rho})$ is empty, then $U = V$ and all minimal lifts of $\overline{\rho}$ will appear in $H^0(X_U, \omega)$. As indicated above, there is a complication if $T(\overline{\rho})$ is non-empty. In this case, minimal modular lifts of $\overline{\rho}$ will appear in the $\sigma^* := \mathrm{Hom}(\sigma, \mathcal{O})$-isotypical part of $H^0(X_V, \omega)$).

For each prime $x \in S(\overline{\rho})$, let c_x denote the Artin conductor of $\overline{\rho}|G_x$. Note that c_x is even when $x \in T(\overline{\rho})$. For $x \in S(\overline{\rho})$, we define subgroups $V_x \subset U_x \subset \mathrm{GL}_2(\mathbf{Z}_x)$ as follows:

- If $x \in P(\overline{\rho})$, we let

$$U_x = V_x = \left\{ g \in \mathrm{GL}_2(\mathbf{Z}_x) : g \equiv \begin{pmatrix} * & * \\ 0 & d \end{pmatrix} \right.$$
$$\left. \mod x^{c_x},\ d \in (\mathbf{Z}/x^{c_x})^\times \text{ has } p-\text{power order} \right\}.$$

- If $x \in T(\overline{\rho})$, then let $U_x = \mathrm{GL}_2(\mathbf{Z}_x)$ and

$$V_x = \ker\left(\mathrm{GL}_2(\mathbf{Z}_x) \longrightarrow \mathrm{GL}_2(\mathbf{Z}/x^{c_x/2})\right).$$

- If $x \in S(\overline{\rho}) - (T(\overline{\rho}) \cup P(\overline{\rho}))$,

$$U_x = V_x = \left\{ g \in \mathrm{GL}_2(\mathbf{Z}_x) : g \equiv \begin{pmatrix} * & * \\ 0 & 1 \end{pmatrix} \mod x^{c_x} \right\}.$$

For x a prime not in $S(\overline{\rho})$, we let

$$U_x = V_x = \mathrm{GL}_2(\mathbf{Z}_x)$$

Finally, if x is any rational prime, we define subgroups $U_{1,x} \subset U_{0,x} \subset \mathrm{GL}_2(\mathbf{Z}_x)$ by:

$$U_{0,x} = \left\{ g \in \mathrm{GL}_2(\mathbf{Z}_x) : g \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \mod x \right\}$$
$$U_{1,x} = \left\{ g \in \mathrm{GL}_2(\mathbf{Z}_x) : g \equiv \begin{pmatrix} * & * \\ 0 & 1 \end{pmatrix} \mod x \right\}.$$

We now set

$$U = \prod_x U_x, \qquad U_i(Q) = \prod_{x \notin Q} U_x \times \prod_{x \in Q} U_{i,x}$$
$$V = \prod_x V_x, \qquad V_i(Q) = \prod_{x \notin Q} V_x \times \prod_{x \in Q} U_{i,x},$$

for $i = 0, 1$. For W equal to one of U, V, $U_i(Q)$ or $V_i(Q)$, we have a smooth projective modular curve X_W over $\mathrm{Spec}(\mathcal{O})$ which is a moduli space of generalized elliptic curves with W-level structure[4]. Let $Y_W \subset X_W$ be the open curve parametrizing genuine elliptic curves and let $j : Y_W \hookrightarrow X_W$ denote the

[4] Again, in order to obtain a representable moduli problem, we may need to introduce auxiliary level structure at a prime q as in Sect. 3.3. This would be necessary if every prime in $S(\overline{\rho})$ were vexing, for example.

inclusion. As in Sect. 3.2, we let $\pi : \mathcal{E} \to X_W$ denote the universal generalized elliptic curve, we let $\omega := \pi_* \omega_{\mathcal{E}/X_W}$ and we let ∞ denote the reduced divisor supported on the cusps. If M is an $\mathcal{O}$-module and $\mathcal{L}$ is a sheaf of $\mathcal{O}$-modules on X_M, then we denote by $\mathcal{L}_M$ the sheaf $\mathcal{L} \otimes_{\mathcal{O}} M$ on X_W. If R is an $\mathcal{O}$-algebra, we will sometimes denote $X_W \times_{\mathrm{Spec}(\mathcal{O})} \mathrm{Spec}(R)$ by $X_{W,R}$.

There is a natural right action of U/V on X_V coming from the description of X_V as a moduli space of generalized elliptic curves with level structure [42, Sect. IV]. It follows from [42, IV 3.10] that we have $X_V/(U/V) \xrightarrow{\sim} X_U$. Away from the cusps, the map $Y_V \to Y_U$ is étale and Galois with Galois group U/V and the map $X_V \to X_U$ is tamely ramified. Similar remarks apply to the maps $X_{V_i(Q)} \to X_{U_i(Q)}$ for $i = 0, 1$.

The natural map $X_{U_1(Q)} \to X_{U_0(Q)}$ is étale and Galois with Galois group

$$\Delta_Q := \prod_{x \in Q} U_{0,x}/U_{1,x} \cong \prod_{x \in Q} (\mathbf{Z}/x)^\times.$$

3.9.1 *Cutting out spaces of modular forms*

Let $G = U/V = \prod_{x \in T} \mathrm{GL}_2(\mathbf{Z}/x^{c_x/2})$ and let σ denote a representation of G on a finite free $\mathcal{O}$-module W_σ. We will now proceed to define a vector bundle $\mathcal{L}_\sigma$ on X such that

$$\begin{aligned} H^0\left(X_U, \omega^{\otimes n} \otimes_{\mathcal{O}_X} \mathcal{L}_\sigma\right) &\xrightarrow{\sim} \left(H^0\left(X_V, \omega^{\otimes n}\right) \otimes_{\mathcal{O}} W_\sigma\right)^G \\ &= \mathrm{Hom}_{\mathcal{O}[G]}\left(W_\sigma^*, H^0(X_V, \omega^{\otimes n})\right), \end{aligned}$$

where W_σ^* is the $\mathcal{O}$-dual of W_σ. The sheaf $\mathcal{L}_\sigma$ will thus allow us to extract the W_σ^*-part of the space of modular forms at level V. We shall also define a cuspidal version $\mathcal{L}_\sigma^{\mathrm{sub}} \subset \mathcal{L}_\sigma$ which extracts the W_σ^*-part of the space of cusp forms at level V:

$$\begin{aligned} H^0(X_U, \omega^{\otimes n} \otimes_{\mathcal{O}_X} \mathcal{L}_\sigma^{\mathrm{sub}}) &\xrightarrow{\sim} (H^0(X_V, \omega^{\otimes n}(-\infty)) \otimes_{\mathcal{O}} W_\sigma)^G \\ &= \mathrm{Hom}_{\mathcal{O}[G]}(W_\sigma^*, H^0(X_V, \omega^{\otimes n}(-\infty))). \end{aligned}$$

The definitions are as follows. Let f denote the natural map $X_V \to X_U$ and define

$$\begin{aligned} \mathcal{L}_\sigma &:= (f_*(\mathcal{O}_{X_V} \otimes_{\mathcal{O}} W_\sigma))^G \\ \mathcal{L}_\sigma^{\mathrm{sub}} &:= (f_*(\mathcal{O}_{X_V}(-\infty) \otimes_{\mathcal{O}} W_\sigma))^G, \end{aligned}$$

where G acts diagonally in both cases. Note that

$$\mathcal{L}_\sigma = (f_* f^*(\mathcal{O}_{X_U} \otimes_{\mathcal{O}} W_\sigma))^G = ((f_* \mathcal{O}_{X_V}) \otimes_{\mathcal{O}} W_\sigma)^G$$

by the projection formula. Similarly, by arguing locally, we see that

$$\mathcal{L}_\sigma^{\mathrm{sub}} = (f_*(\mathcal{O}_{X_V}(-\infty)) \otimes_{\mathcal{O}} W_\sigma)^G.$$

For $i = 0, 1$ we denote the pull back of $\mathcal{L}_\sigma$ to $X_{U_i(Q)}$ also by $\mathcal{L}_\sigma$. This notation is justified since, by flatness of the map $X_{U_i(Q)} \to X_U$, the pullback is isomorphic to $(f_*(\mathcal{O}_{X_{V_i(Q)}} \otimes_{\mathcal{O}} W_\sigma))^G$, where we continue to denote by f the natural map $X_{V_i(Q)} \to X_{U_i(Q)}$. When we use the same notation for sheaves on different spaces, the underlying spaces will always be clear from the context.

On $X_{U_i(Q)}$ we reserve the notation $\mathcal{L}_\sigma^{\mathrm{sub}}$ for $(f_*(\mathcal{O}_{X_{V_i(Q)}}(-\infty)) \otimes_{\mathcal{O}} W_\sigma)^G$. The pull back of $\mathcal{L}_\sigma^{\mathrm{sub}}$ on X_U to $X_{U_i(Q)}$ is a sub-sheaf of $\mathcal{L}_\sigma^{\mathrm{sub}}$ on $X_{U_i(Q)}$ (the quotient being supported at ramified cusps).

We now discuss Hecke actions on cohomology. Let X denote $X_{U_i(Q)}$ and let $X(V)$ denote $X_{V_i(Q)}$ for some choice of $i = 0$ or 1. Let f denote the map $X(V) \to X$. (Note that if Q is empty, then we recover $f : X_V \to X_U$). Let $x \notin S(\overline{\rho}) \cup \{p\}$. As in Sect. 3.2.3, we have a modular curve $X_0(x)$, obtained from X by the addition of an appropriate level structure at x, together with degeneracy maps $\pi_1, \pi_2 : X_0(x) \to X$. (The level structure at x depends on whether or not $x \in Q$). We define $X_0(V; x)$ similarly, starting from $X(V)$. The natural map $X_0(V; x) \to X_0(x)$ is again denoted f. Then note that we have a natural isomorphism

$$\phi(\sigma)_{12} : \pi_2^* \mathcal{L}_\sigma \xrightarrow{\sim} \pi_1^* \mathcal{L}_\sigma$$

of sheaves on $X_0(x)$. Indeed for $i = 0, 1$, by flatness of the map $\pi_i : X_0(x) \to X$, the pullback $\pi_i^* \mathcal{L}_\sigma$ is canonically isomorphic to

$$(f_*(\mathcal{O}_{X_0(V;x)} \otimes_{\mathcal{O}} W_\sigma))^G,$$

independently of i. (The only point to note is that the morphism $X_0(x) \times_{\pi_i, X} X(V) \to X_0(x)$ is canonically isomorphic to $X_0(V; x) \to X_0$). Similarly, if $a \in \mathbf{Z}$ is coprime to the elements of $S(\overline{\rho}) \cup Q$, we have a morphism $\langle a \rangle : X \to X$ which corresponds to multiplication by a on the level structure. Then $\langle a \rangle^* \mathcal{L}_\sigma$ is canonically isomorphic to $\mathcal{L}_\sigma$.

Let M denote an $\mathcal{O}$-module and let n be an integer. Then using the isomorphisms $\pi_2^* \mathcal{L}_\sigma \xrightarrow{\sim} \pi_1^* \mathcal{L}_\sigma$ of the previous paragraph, and following the definitions of Sect. 3.2.3, we can define Hecke operators on the cohomology of $\omega^n \otimes \mathcal{L}_\sigma \otimes_{\mathcal{O}} M$. For example, $x T_x$ is defined as the composite (taking $M = \mathcal{O}$ for simplicity):

$$\begin{aligned} H^i(X_U, \omega^n \otimes \mathcal{L}_\sigma) &\xrightarrow{\pi_2^*} H^i(X_0(U;x), \pi_2^*\omega^n \otimes \mathcal{L}_\sigma) \\ &\xrightarrow{\phi_{12}\otimes\phi_{12}(\sigma)} H^i(X_0(U;x), \pi_1^*\omega^n \otimes \mathcal{L}_\sigma) \\ &\xrightarrow{\mathrm{tr}(\pi_1)} H^i(X_U, \omega^n \otimes \mathcal{L}_\sigma). \end{aligned}$$

fLet $\mathcal{L}_\sigma^{\mathrm{sub}}$ denote the sheaf

$$(f_*(\mathcal{O}_{X_0(V;x)} \otimes_{\mathcal{O}} W_\sigma))^G,$$

on $X_0(U;x)$. (In what follows we will be using $\mathcal{L}_\sigma^{\mathrm{sub}}$ and $\mathcal{L}_\sigma$ to denote sheaves on both X_U and $X_0(U;x)$, but the underlying space will be clear in each instance). We then have canonical inclusions

$$\pi_1^*(\mathcal{L}_\sigma^{\mathrm{sub}}), \pi_2^*(\mathcal{L}_\sigma^{\mathrm{sub}}) \subset \mathcal{L}_\sigma^{\mathrm{sub}}$$

of sheaves on $X_0(U;x)$. Note also that the composition of morphisms of sheaves on X_U

$$\pi_{1,*}(\mathcal{L}_\sigma^{\mathrm{sub}}) \hookrightarrow \pi_{1,*}(\mathcal{L}_\sigma) = \pi_{1,*}(\pi_1^*\mathcal{L}_\sigma) \xrightarrow{\mathrm{tr}(\pi_1)} \mathcal{L}_\sigma$$

factors through the sheaf $\mathcal{L}_\sigma^{\mathrm{sub}}$. This then allows us to define Hecke operators on the cohomology of $\omega^n \otimes \mathcal{L}_\sigma^{\mathrm{sub}} \otimes_{\mathcal{O}} M$.

In summary, we have operators:

- T_x and $\langle a \rangle$ on

$$H^j(X_U, \omega^n \otimes_{\mathcal{O}_{X_U}} \mathcal{L}_\sigma^{\mathrm{sub}} \otimes_{\mathcal{O}} M) \text{ and } H^j(X_U, \omega^n \otimes_{\mathcal{O}_{X_U}} \mathcal{L}_\sigma \otimes_{\mathcal{O}} M)$$

  for all $x \notin S(\overline{\rho}) \cup \{p\}$ and a coprime to the elements of $S(\overline{\rho})$, and
- $T_x, U_y, \langle a \rangle$ on

$$H^j(X_{U_i(Q)}, \omega^n \otimes_{\mathcal{O}_{X_U}} \mathcal{L}_\sigma^{\mathrm{sub}} \otimes_{\mathcal{O}} M) \text{ and } H^j(X_{U_i(Q)}, \omega^n \otimes_{\mathcal{O}_{X_U}} \mathcal{L}_\sigma \otimes_{\mathcal{O}} M)$$

  for all $x \notin S(\overline{\rho}) \cup Q \cup \{p\}$, $y \in Q$ and a coprime to the elements of $S(\overline{\rho}) \cup Q$.

Part (2) of the following lemma shows that $\mathcal{L}_\sigma$ and $\mathcal{L}_\sigma^{\mathrm{sub}}$ do indeed allow us to extract the W_σ^*-part of the space of modular forms at level V.

Lemma 3.27 *Let X denote X_U (resp. $X_{U_i(Q)}$ for $i = 0$ or 1), let $X(V)$ denote X_V (resp. $X_{V_i(Q)}$) and let f denote the map $X(V) \to X$. Then*

(1) *The sheaves* $\mathcal{L}_\sigma^{\mathrm{sub}}$ *and* $\mathcal{L}_\sigma$ *are locally free of finite rank on* X.
(2) *If* A *is an* $\mathcal{O}$*-algebra and* $\mathcal{V}$ *is a coherent locally free sheaf of* $\mathcal{O}_{X_A}$*-modules, then*

$$H^0(X_A, \mathcal{V} \otimes_{\mathcal{O}_{X_A}} (\mathcal{L}_\sigma)_A) \xrightarrow{\sim} (H^0(X(V)_A, f^*\mathcal{V}) \otimes_{\mathcal{O}} W_\sigma)^G$$
$$H^0(X_A, \mathcal{V} \otimes_{\mathcal{O}_{X_A}} (\mathcal{L}_\sigma^{\mathrm{sub}})_A) \xrightarrow{\sim} (H^0(X(V)_A, (f^*\mathcal{V})(-\infty)) \otimes_{\mathcal{O}} W_\sigma)^G.$$

Proof We give the proof for $\mathcal{L}_\sigma$; the case of $\mathcal{L}_\sigma^{\mathrm{sub}}$ is treated in exactly the same way. Let Y (resp. $Y(V)$) denote the non-cuspidal open subscheme of X (resp. $X(V)$). We have $\mathcal{L}_\sigma|_Y = f_*(\mathcal{O}_{Y(V)} \otimes_{\mathcal{O}} W_\sigma)^G$ since the inclusion $Y \to X$ is flat. Since the map $Y(V) \to Y$ is étale, it follows from [43, Sect. III.12 Theorem 1 (B)] (and its proof) that $\mathcal{L}_\sigma|_Y \xrightarrow{\sim} \mathcal{O}_Y \otimes_{\mathcal{O}} W_\sigma$. To show that $\mathcal{L}_\sigma$ is locally free of finite rank on X, it remains to check that its stalks at points of $X - Y$ are free. Let x be a point of $X - Y$. We can and do assume that for each point x' of $X(V)$ lying above x, the natural map on residue fields is an isomorphism. We have

$$\mathcal{L}_{\sigma,x} = \left(\bigoplus_{x' \mapsto x} \mathcal{O}_{X_V,x'} \otimes W_\sigma \right)^G.$$

Choose some point $x' \mapsto x$ and let $I(x'/x) \subset G$ be the inertia group of x'. Then projection onto the x'-component defines an isomorphism

$$\left(\bigoplus_{x'' \mapsto x} \mathcal{O}_{X(V),x''} \otimes W_\sigma \right)^G \xrightarrow{\sim} \left(\mathcal{O}_{X(V),x'} \otimes W_\sigma\right)^{I(x'/x)}.$$

Now $I(x'/x)$ is abelian of order prime to p (see [42, Sect. VI.5]). Extending $\mathcal{O}$, we may assume that each character χ of $I(x'/x)$ is defined over $\mathcal{O}$. Let $W_{\sigma,\chi}$ and $\mathcal{O}_{X_V,x',\chi}$ denote the χ-parts of W_σ and $\mathcal{O}_{X_V,x'}$. Then $W_\sigma \cong \oplus_\chi W_{\sigma,\chi}$ and similarly $\mathcal{O}_{X(V),x'} \cong \oplus_\chi \mathcal{O}_{X(V),x',\chi}$. Each $W_{\sigma,\chi}$ (resp. $\mathcal{O}_{X(V),x',\chi}$) is free over $\mathcal{O}$ (resp. $\mathcal{O}_{X,x}$), being a summand of a free module. (Note that f is finite flat). We now have

$$\mathcal{L}_{\sigma,x} \xrightarrow{\sim} \left(\mathcal{O}_{X(V),x'} \otimes W_\sigma\right)^{I(x'/x)} \xrightarrow{\sim} \bigoplus_\chi W_{\sigma,\chi} \otimes_{\mathcal{O}} \mathcal{O}_{X(V),x',\chi^{-1}},$$

which is free over $\mathcal{O}_{X,x}$. This establishes part (1).

We now turn to part (2). We first of all note that the proof of the previous part shows that

$$(\mathcal{L}_\sigma)_A \xrightarrow{\sim} (f_*(\mathcal{O}_{X(V)_A} \otimes_{\mathcal{O}} W_\sigma))^G,$$

as sheaves on X_A. Now, let $\mathcal{V}$ be as in the statement of the lemma. Then,

$$\begin{aligned}\mathcal{V} \otimes_{\mathcal{O}_{X_A}} (\mathcal{L}_\sigma)_A &\cong \mathcal{V} \otimes_{\mathcal{O}_{X_A}} (f_*(\mathcal{O}_{X(V)_A} \otimes_{\mathcal{O}} W_\sigma))^G \\ &\cong (\mathcal{V} \otimes_{\mathcal{O}_{X_A}} f_*(\mathcal{O}_{X(V)_A} \otimes_{\mathcal{O}} W_\sigma))^G \\ &\cong (f_* f^*(\mathcal{V} \otimes_{\mathcal{O}} W_\sigma))^G.\end{aligned}$$

Here G acts trivially on $\mathcal{V}$ and the third isomorphism follows from the projection formula. Taking global sections we obtain,

$$\begin{aligned}H^0(X_A, \mathcal{V} \otimes (\mathcal{L}_\sigma)_A) &= (H^0(X(V)_A, f^*(\mathcal{V} \otimes_{\mathcal{O}} W_\sigma))^G \\ &= (H^0(X(V)_A, f^*(\mathcal{V})) \otimes_{\mathcal{O}} W_\sigma)^G,\end{aligned}$$

as required. □

Let X and $X(V)$ be as in the statement of the previous lemma. Let $\sigma^* = \mathcal{H}om_{\mathcal{O}}(W_\sigma, \mathcal{O})$ be the dual of the representation σ. We now consider the dual vector bundle $\mathcal{L}_\sigma^* = \mathcal{H}om_{\mathcal{O}_X}(\mathcal{L}_\sigma, \mathcal{O}_X)$ and its relation to $\mathcal{L}_{\sigma^*}$. In addition, we let A denote an $\mathcal{O}$-algebra and we consider the situation base changed to $\operatorname{Spec} A$. First of all, note that we have an isomorphism

$$X(V)_A/G \xrightarrow{\sim} X_A$$

and in particular, $\mathcal{O}_{X_A} \xrightarrow{\sim} f_*(\mathcal{O}_{X(V)_A})^G$. (When $A = \mathcal{O}$, this follows from [42, Sect. IV Proposition 3.10]. The same argument works when $A = k$. These two cases, and the flatness of $X(V)$ over $\mathcal{O}$, imply the result when $A = \mathcal{O}/\varpi^n$. The general result follows from this by [44, Proposition A7.1.4]. Alternatively, as pointed out to us by the referee, one can see directly that $X(V)_A/G = X_A$ by applying the argument of the proof of Lemma 3.27 (1)). We have shown in the proof of Lemma 3.27 (2) that

$$(\mathcal{L}_\sigma)_A \cong (f_*(\mathcal{O}_{X(V)_A} \otimes_{\mathcal{O}} W_\sigma))^G,$$

as sheaves on X_A. By the projection formula, we therefore also have

$$(\mathcal{L}_\sigma)_A \cong (f_*(\mathcal{O}_{X(V)_A}) \otimes_{\mathcal{O}} W_\sigma)^G.$$

Applying this with σ^* in place of σ, we see that

$$\begin{aligned}(\mathcal{L}_{\sigma^*})_A &\cong (f_*(\mathcal{O}_{X(V)_A}) \otimes_{\mathcal{O}} \mathcal{H}om_{\mathcal{O}}(W_\sigma, \mathcal{O}))^G \\ &\cong \left(\mathcal{H}om_{f_*(\mathcal{O}_{X(V)_A})}(f_*(\mathcal{O}_{X(V)_A}) \otimes_{\mathcal{O}} W_\sigma, f_*(\mathcal{O}_{X(V)_A}))\right)^G\end{aligned}$$

Since $\mathcal{O}_{X_A} \xrightarrow{\sim} f_*(\mathcal{O}_{X(V)_A})^G$, we have a map

$$(\mathcal{L}_{\sigma^*})_A \longrightarrow \mathcal{H}om_{\mathcal{O}_{X_A}}((\mathcal{L}_\sigma)_A, \mathcal{O}_{X_A}) = (\mathcal{L}_\sigma)^*_A$$

given by restriction to G-invariants. This map is induced from the corresponding map $\mathcal{L}_{\sigma^*} \to \mathcal{L}^*_\sigma$ over $\mathrm{Spec}\mathcal{O}$. In addition, when restricted to Y_A, this map is an isomorphism since the equivalence of [43, Sect. III.12 Theorem 1 (B)] for locally free sheaves is compatible with taking duals. In particular, the map $\mathcal{L}_{\sigma^*} \to \mathcal{L}^*_\sigma$ is injective and remains injective after base change to $\mathrm{Spec}A$, for all A.

Lemma 3.28 *The injection $\mathcal{L}_{\sigma^*} \hookrightarrow \mathcal{L}^*_\sigma$ restricts to an isomorphism*

$$\mathcal{L}^{\mathrm{sub}}_{\sigma^*} \xrightarrow{\sim} (\mathcal{L}^*_\sigma)(-\infty).$$

Similarly, we have

$$(\mathcal{L}^{\mathrm{sub}}_\sigma)^*(-\infty) \cong \mathcal{L}_{\sigma^*}.$$

Proof The second statement follows immediately from the first by reversing the roles of σ and σ^*. Thus, we consider the first statement. Away from the cusps, all three inclusions

$$\mathcal{L}^{\mathrm{sub}}_{\sigma^*} \hookrightarrow \mathcal{L}_{\sigma^*} \hookrightarrow \mathcal{L}^*_\sigma \quad \text{and} \quad (\mathcal{L}^*_\sigma)(-\infty) \hookrightarrow \mathcal{L}^*_\sigma$$

are isomorphisms. It therefore suffices to show that at each closed point x of $\mathcal{C}$, the natural map gives rise to an isomorphism

$$(\mathcal{L}^{\mathrm{sub}}_{\sigma^*})^\wedge_x \xrightarrow{\sim} (\mathcal{L}^*_\sigma)(-\infty)^\wedge_x$$

along the formal completions at x. Extending $\mathcal{O}$ if necessary, we may assume that all cusps of X and $X(V)$ are defined over $\mathcal{O}$ and for each point x' of $X(V)$ lying over x, there is a uniformizer q at x' so the map

$$\mathcal{O}^\wedge_{X,x} \to \mathcal{O}^\wedge_{X(V),x'}$$

is isomorphic to

$$\mathcal{O}[q^e] \to \mathcal{O}[q].$$

Here, $e = \#I(x'/x)$ and we may assume that $\mathcal{O}$ contains the primitive e-th roots of unity and the inertia group $I(x'/x)$ is isomorphic to $\mu_e \subset \mathcal{O}^\times$ via $\sigma \mapsto \frac{\sigma(q)}{q}$. Choose a primitive e-th root of unity ζ and for $i = 0, \ldots, e-1$, let $\chi_i : I(x'/x) \cong \mu_e \to \mathcal{O}^\times$ be the character which sends ζ to ζ^i. Then, the χ_i-part of $\mathcal{O}^\wedge_{X(V),x'}$ is given by

$$\mathcal{O}^\wedge_{X(V),x',\chi_i} = q^i \mathcal{O}[q^e] \subset \mathcal{O}[q].$$

Thus, as in the proof of Lemma 3.27, we have

$$(\mathcal{L}_\sigma)^\wedge_x \cong \bigoplus_{i=0}^{e-1} q^i \mathcal{O}[q^e] \otimes_{\mathcal{O}} W_{\sigma,\chi_i^{-1}}.$$

Taking duals over $\mathcal{O}^\wedge_{X,x} \cong \mathcal{O}[q^e]$, we obtain

$$(\mathcal{L}^*_\sigma)^\wedge_x \cong \bigoplus_{i=0}^{e-1} q^{-i} \mathcal{O}[q^e] \otimes_{\mathcal{O}} (\mathrm{Hom}_{\mathcal{O}}(W_{\sigma,\chi_i^{-1}}, \mathcal{O})).$$

Note that $\mathrm{Hom}_{\mathcal{O}}(W_{\sigma,\chi_i^{-1}}, \mathcal{O}) = W_{\sigma^*,\chi_i}$ and since q^e is a uniformizer at x, we obtain, under the natural map, identifications

$$\begin{aligned}(\mathcal{L}^*_\sigma)(-\infty)^\wedge_x &\cong \bigoplus_{i=0}^{e-1} q^{e-i} \mathcal{O}[q^e] \otimes_{\mathcal{O}} W_{\sigma^*,\chi_i} \\ &\cong \bigoplus_{i=1}^{e} q^{i} \mathcal{O}[q^e] \otimes_{\mathcal{O}} W_{\sigma^*,\chi_i^{-1}}.\end{aligned}$$

This is precisely $(\mathcal{L}^{\mathrm{sub}}_{\sigma^*})^\wedge_x$ by the proof of Lemma 3.27 (1). □

We deduce the following.

Corollary 3.29 *If $n > 1$, then*

$$H^1(X, \omega^{\otimes n} \otimes \mathcal{L}_\sigma) = \{0\}$$

and hence

$$H^0(X, \omega^{\otimes n} \otimes \mathcal{L}_\sigma \otimes_{\mathcal{O}} \mathcal{O}/\varpi^m) = (H^0(X(V), \omega^{\otimes n}) \otimes_{\mathcal{O}} W_\sigma)^G \otimes_{\mathcal{O}} \mathcal{O}/\varpi^m.$$

Moreover, the analogous result holds for $n > 2$ if we replace $\mathcal{L}_\sigma$ by $\mathcal{L}^{\mathrm{sub}}_\sigma$.

Proof The second statement follows immediately from the first and from Lemma 3.27 (2) by considering the long exact sequence in cohomology associated to the short exact sequence

$$0 \longrightarrow \omega^{\otimes n} \otimes_{\mathcal{O}_X} \mathcal{L}_\sigma \xrightarrow{\varpi^m} \omega^{\otimes n} \otimes_{\mathcal{O}_X} \mathcal{L}_\sigma \longrightarrow (\omega^{\otimes n} \otimes_{\mathcal{O}_X} \mathcal{L}_\sigma)/\varpi^m \longrightarrow 0.$$

To prove the first statement, it suffices to show that $H^1(X, \omega^{\otimes n} \otimes_{\mathcal{O}_X} \mathcal{L}_\sigma \otimes_{\mathcal{O}} k) = \{0\}$. By Serre duality, this is equivalent to the vanishing of $H^0(X_k, \omega^{\otimes(2-n)} \otimes_{\mathcal{O}_X} \mathcal{L}^*_\sigma(-\infty))$. By Lemma 3.28, we are therefore reduced to showing $H^0(X_k, \omega^{\otimes(2-n)} \otimes_{\mathcal{O}_X} \mathcal{L}^{\mathrm{sub}}_{\sigma^*}) = 0$. However, by Lemma 3.27 (2) again, we have

$$H^0(X_k, \omega^{\otimes(2-n)} \otimes_{\mathcal{O}_X} \mathcal{L}^{\mathrm{sub}}_{\sigma^*}) \cong (H^0(X(V)_k, \omega^{\otimes(2-n)}(-\infty)) \otimes_{\mathcal{O}} W_{\sigma^*})^G$$

which vanishes since $n > 1$. The case of $\mathcal{L}^{\mathrm{sub}}_\sigma$ is proved in the same way. □

3.9.2 The proof of Theorem 1.4 in the presence of vexing primes

To complete the proof of Theorem 1.4, it suffices to note the various modifications which must be made to the argument. For vexing primes x, let c_x denote the conductor of $\overline{\rho}$ (which is necessarily even). We define a $\mathcal{O}$-representation W_{σ_x} of $\mathrm{GL}_2(\mathbf{Z}/x^{c_x/2}\mathbf{Z})$ to be the representation σ_x as in Sect. 5 of [15]. The collection $\sigma = (\sigma_x)_{x \in T(\overline{\rho})}$ gives rise to a sheaf $\mathcal{L}_\sigma$ on X_U as above. Let $\mathbf{T}_\emptyset$ denote the ring of Hecke operators acting on $H_0(X_U, \omega \otimes \mathcal{L}_\sigma)$ generated by Hecke operators away from $S(\overline{\rho}) \cup \{p\}$. The analogue of Theorem 3.26 is as follows:

Theorem 3.30 *The map $\varphi : R^{\min} \twoheadrightarrow \mathbf{T}_{\emptyset,\mathfrak{m}_\emptyset}$ is an isomorphism and $\mathbf{T}_{\emptyset,\mathfrak{m}_\emptyset}$ acts freely on $H_0(X_U, \omega \otimes \mathcal{L}_\sigma)_{\mathfrak{m}_\emptyset}$.*

Proof The proof is the same as the proof of Theorem 3.26; we indicate below the modifications that need to be made.

(1) (Lemma 3.5): Exactly the same argument shows that there is an isomorphism of Hecke modules:

$$H^0(X_U, (\omega \otimes \mathcal{L}_\sigma)_{K/\mathcal{O}})_{\mathfrak{m}_\emptyset} \xrightarrow{\sim} H^0(X_{U_0(Q)}, (\omega \otimes \mathcal{L}_\sigma)_{K/\mathcal{O}})_{\mathfrak{m}}.$$

The only point to note is that there is an operator

$$W_x : H^0(X_{U_0(x)}, (\omega \otimes \mathcal{L}_\sigma)_{K/\mathcal{O}}) \to H^0(X_{U_0(x)}, (\omega \otimes \mathcal{L}_\sigma)_{K/\mathcal{O}})$$

such that $W_x^2 = x\langle x\rangle$ and

$$\frac{1}{x}\pi_1^* \circ \mathrm{tr}(\pi_1) \circ W_x = U_x + \frac{1}{x} W_x.$$

To see this, one can note that the corresponding operator W_x on $H^0(X_{V_0(x)}, \omega_{K/\mathcal{O}})$ (defined in Sect. 3.2.3) commutes with the action of $G = U/V$, and hence induces the desired operator W_x on

$$H^0(X_{U_0(x)}, (\omega \otimes \mathcal{L}_\sigma)_{K/\mathcal{O}}) = (H^0(X_{V_0(x)}, \omega_{K/\mathcal{O}}) \otimes_{\mathcal{O}} W_\sigma)^G.$$

(2) (Proposition 3.7 (2)): The corresponding statement holds: namely, the natural map

$$H^i(X_{U_\Delta(Q)}, (\omega \otimes \mathcal{L}_\sigma^{\mathrm{sub}})_{K/\mathcal{O}})_{\mathfrak{m}} \to H^i(X_{U_\Delta(Q)}, (\omega \otimes \mathcal{L}_\sigma)_{K/\mathcal{O}})_{\mathfrak{m}}$$

is an isomorphism for $i = 0, 1$ whenever $\mathfrak{m}$ is non-Eistenstein. Indeed, if $\mathcal{C}_{V_\Delta(Q)} \subset X_{V_\Delta(Q)}$ is the cuspidal subscheme, then we have an exact sequence of sheaves on $X_{U_\Delta(Q)}$:

$$0 \longrightarrow \omega \otimes \mathcal{L}_\sigma^{\mathrm{sub}} \longrightarrow \omega \otimes \mathcal{L}_\sigma \longrightarrow (f_*((\omega \otimes_{\mathcal{O}} W_\sigma)|_{\mathcal{C}_{V_\Delta(Q)}}))^G,$$

and it suffices to show that the cohomology of the last term (which is concentrated in degree 0) is Eisenstein. However, the argument of Remark 3.4 (noting that the group $\prod_{x \in T(\overline{\rho})} \mathrm{GL}_2(\mathbf{Z}_x) \times \prod_{x \in P(\overline{\rho})} \mathbf{Z}_x^\times$ acts transitively on the set of cusps in $X_{V_\Delta(Q)}$) shows that

$$H^0(X_{U_\Delta(Q)}, (f_*((\omega \otimes W_\sigma)|_{\mathcal{C}_{V_\Delta(Q)}}))^G) = (H^0(\mathcal{C}_{V_\Delta(Q)}, \omega) \otimes_{\mathcal{O}} W_\sigma)^G$$

is Eisenstein, which gives the desired result.

(3) (Proposition 3.8): We need to show that the $\mathcal{O}[\Delta]$-module $M = H_0(X_\Delta(Q), \omega \otimes \mathcal{L}_\sigma)_{\mathfrak{m}}$ is balanced. First of all note that $\Omega^1_{X_\Delta(Q)/\mathcal{O}} \otimes \mathcal{L}_\sigma^* = \omega^2(-\infty) \otimes \mathcal{L}_\sigma^* = \omega^2 \otimes \mathcal{L}_{\sigma^*}^{\mathrm{sub}}$ by Lemma 3.28, and hence

$$H_i(X_{U_\Delta(Q)}, \omega^n \otimes \mathcal{L}_\sigma) = H^i(X_{U_\Delta(Q)}, (\omega^{2-n} \otimes \mathcal{L}_{\sigma^*}^{\mathrm{sub}})_{K/\mathcal{O}})^\vee.$$

We use this to endow the left hand side with a Hecke action.
We now modify definitions of Φ and Ψ from Sect. 3.2.4: let Φ denote the composition of isomorphisms:

$$H^1(X_{U_\Delta(Q)}, (\omega^{2-n} \otimes \mathcal{L}_{\sigma^*}^{\mathrm{sub}})_{K/\mathcal{O}}) \xrightarrow{D} H^0(X_{U_\Delta(Q)}, \Omega \otimes \omega^{n-2} \otimes \mathcal{L}_\sigma(\infty))^\vee \xrightarrow{KS^\vee} H^0(X_{U_\Delta(Q)}, \omega^n \otimes \mathcal{L}_\sigma)^\vee,$$

where D is Verdier duality, and KS is the Kodaira–Spencer isomorphism, and we have used Lemma 3.28. Then by the proof of [31, Proposition 7.3], we have:

$$\Phi \circ T_x = x^{1-n} T_x^{t,\vee} \circ \Phi,$$

for all x prime to NQ, and the same relation holds for the operators U_y with $y|Q$. We also have $\Phi \circ \langle a \rangle = \langle a^{-1} \rangle \circ \Phi$ for $x|NQ$ because D switches $\langle a \rangle^*$

with $\langle a\rangle_*^\vee = \langle a^{-1}\rangle^{*,\vee}$. The transposed operators T_x^t and U_x^t are defined in [31]. We have:

$$(T_x^t f)(E, \alpha_{NQ}) = \sum_{\phi:(E',\alpha'_{NQ})\to(E,\alpha_{NQ})} \phi^{t,*} f(E', \alpha'_{NQ})$$

where the sum is over all x-isogenies ϕ such that $\phi \circ \alpha'_{NQ} = \alpha_{NQ}$. In the definition of T_x^t, we can replace the sum over ϕ by the sum over the corresponding dual isogenies $E \to E'$. However, note that if $\phi : (E', \alpha'_{NQ}) \to (E, \alpha_{NQ})$ is compatible with level structures at NQ, then so is $\phi^t : (E, \alpha_{NQ}) \to (E', x \circ \alpha'_{NQ})$. In this way we see that

$$T_x^t = \langle x^{-1}\rangle T_x$$

on $H^0(X_{U_\Delta(Q)}, \omega^n)$. The Pontryagin dual $\Psi := \Phi^\vee$ is thus an isomorphism

$$\Psi : H^0(X_{U_\Delta(Q)}, \omega^n \otimes \mathcal{L}_\sigma) \longrightarrow H_1(X_{U_\Delta(Q)}, \omega^n \otimes \mathcal{L}_\sigma)$$

such that

$$\begin{aligned} \Psi \circ (x^{1-n}\langle x^{-1}\rangle T_x) &= T_x \circ \Psi \\ \Psi \circ (y^{1-n} U_y^t) &= U_y \circ \Psi \\ \Psi \circ \langle a^{-1}\rangle &= \langle a\rangle \circ \Psi. \end{aligned}$$

Now, with $\mathcal{L} = \mathcal{L}_\sigma$, the proof of Proposition 3.8 proceeds in exactly the same manner, up to the point where it suffices to show that

$$\dim_K H_0(X_{U_0(Q),K}, \omega \otimes \mathcal{L}_\sigma)_\mathfrak{m} = \dim_K H_1(X_{U_0(Q),K}, \omega \otimes \mathcal{L}_\sigma)_\mathfrak{m}.$$

As before, by definition, the left hand side of this is equal to

$$\dim_K H^0(X_{U_0(Q),K}, \omega \otimes \mathcal{L}_{\sigma^*}^{\mathrm{sub}})_\mathfrak{m},$$

which in turn, by point (2) above, is equal to:

$$\dim_K H^0(X_{U_0(Q),K}, \omega \otimes \mathcal{L}_{\sigma^*})_\mathfrak{m},$$

On the other hand, using the isomorphism Ψ, we see that the right hand side is equal to:

$$\dim_K H^0(X_{U_0(Q)}, \omega^n \otimes \mathcal{L}_\sigma)_{\mathfrak{m}^*}$$

where $\mathfrak{m}^*$ is a maximal ideal of the polynomial ring R generated over $\mathcal{O}$ by the operators T_x, U_y^t, $\langle a\rangle$. Specifically, let $\alpha : \mathbf{T} \to \mathbf{T}/\mathfrak{m} = k$ be the reduction map. Then $\mathfrak{m}^*$ is the kernel of the map $\beta : R \to k$ defined by: $\beta(T_x) = \alpha(\langle x\rangle T_x)$, $\beta(U_y^t) = \alpha(U_y)$, and $\beta(\langle a\rangle) = \alpha(\langle a^{-1}\rangle)$.
Thus, we need to see that $H^0(X_{U_0(Q),K}, \omega\otimes\mathcal{L}_\sigma)_{\mathfrak{m}^*}$ and $H^0(X_{U_0(Q),K}, \omega\otimes \mathcal{L}_{\sigma^*})_{\mathfrak{m}}$ have the same dimension. One way to see this is as follows: after choosing an embedding $K \hookrightarrow \mathbf{C}$, we can identify both sides in terms of automorphic representations of $\mathrm{GL}_2/\mathbf{Q}$ which are limits of discrete series at ∞, unramified outside $S(\overline{\rho}) \cup Q$ and satisfy appropriate local conditions at the primes in $S(\overline{\rho})\cup Q$. The operation which sends each such automorphic representation π to its contragredient then interchanges

$$H^0(X_{U_0(Q),K}, \omega \otimes \mathcal{L}_\sigma)_{\mathfrak{m}^*} \otimes_K \mathbf{C} \text{ and } H^0(X_{U_0(Q),K}, \omega \otimes \mathcal{L}_{\sigma^*})_{\mathfrak{m}} \otimes_K \mathbf{C},$$

from which the result follows.

(4) (Theorem 3.11): The analogue of this theorem is true. Namely, let $\mathbf{T}$ denote the subalgebra of endomorphisms of

$$H^0(X_{U_1(Q)}, (\omega \otimes \mathcal{L}_\sigma)_{K/\mathcal{O}})$$

generated by the operators T_x, U_y and $\langle a\rangle$. For each $x \in Q$, we assume that the Hecke polynomial $X^2 - T_x X + \langle x\rangle$ has distinct roots in $\mathbf{T}_\emptyset/\mathfrak{m}_\emptyset$ and we let α_x be one of these roots. Let $\mathfrak{m}$ be the ideal of $\mathbf{T}$ generated by $\mathfrak{m}_\emptyset$ and $U_x - \alpha_x$ for $x \in Q$. Then there is a Galois representation

$$\rho_Q : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{T}_{\mathfrak{m}})$$

deforming $\overline{\rho}$, unramified away from $S(\overline{\rho}) \cup \{p\}$ and such that Frob_x has trace T_x for all $x \notin S(\overline{\rho}) \cup \{p\}$. Moreover $\rho'_Q := \rho_Q \otimes \eta$, where η is defined as before, is a deformation of $\overline{\rho}$ minimal outside Q.
This is proved as follows: as before it suffices to fix an $m \geq 1$ and work with the quotient $\mathbf{T}_{\mathfrak{m}}/J_m$ of $\mathbf{T}_{\mathfrak{m}}$ acting faithfully on $H^0(X_{U_1(Q)}, (\omega \otimes \mathcal{L}_\sigma)_{\mathcal{O}/\varpi^m})_{\mathfrak{m}}$. Then we have

$$H^0(X_{U_1(Q)}, (\omega \otimes \mathcal{L}_\sigma)_{\mathcal{O}/\varpi^m})_{\mathfrak{m}} \subset H^0(X_{V_1(Q)}, \omega_{\mathcal{O}/\varpi^m})_{\mathfrak{m}} \otimes_{\mathcal{O}} W_\sigma.$$

From this inclusion and the arguments of Sections 3.3–3.7, we immediately deduce the existence of ρ_Q over $\mathbf{T}_{\mathfrak{m}}/J_m$ such that ρ'_Q satisfies all the conditions of Definition 3.1, except possibly for condition (3). More precisely, we construct a deformation over the Hecke algebra of

$$H^0(X_{V_1(Q)}, \omega_{\mathcal{O}/\varpi^m})_{\mathfrak{m}}$$

which satisfies these properties exactly as we did in Sections 3.3–3.7. (The new level structures at the primes in $T(\overline{\rho})$ do not affect the arguments; the only essential difference is that the modular curves are no longer geometrically connected. Thus, in any argument involving q-expansions, one needs to consider q-expansions at a cusp on each connected component instead of at the single cusp ∞. Note, however, that the use of q-expansions was only used for the following two facts: the identity $\phi \circ T_p - U_p \circ \phi = \langle p \rangle V_p$ and the claim that $\theta V_p = 0$, which was used to show that $(\phi, \phi \circ T_p - U_p \circ \phi)$ was injective. On the other hand, the group $\prod_{x \in T(\overline{\rho})} \mathrm{GL}_2(\mathbf{Z}_x)$ acts invertibly on $X_{V_1(Q)}$ and hence also the cohomology group above, acts transitively on the set of connected components, and commutes with the Hecke operators at p. Hence it suffices to check these identities on the component at ∞, where the required conclusions follow from our previous computation). We then use the above inclusion of Hecke modules to deduce the result over the algebra $\mathbf{T}_{\mathfrak{m}}/J_m$. It remains to show that condition (3) of Definition 3.1 holds. For this, we use that fact that multiplication by a high power of a lift of the Hasse invariant of level $X_{V_1(Q)}$ realizes

$$H^0(X_{U_1(Q)}, (\omega \otimes \mathcal{L}_\sigma)_{\mathcal{O}/\varpi^m})_{\mathfrak{m}}$$

as a Hecke equivariant subquotient of

$$(H^0(X_{V_1(Q)}, \omega^n)_{\mathfrak{m}} \otimes_{\mathcal{O}} W_\sigma)^G$$

for some sufficiently large n. (This follows from Corollary 3.29). It therefore suffices to show that the deformation of $\overline{\rho}$ over the Hecke algebra of

$$(H^0(X_{V_1(Q)}, \omega^n)_{\mathfrak{m}} \otimes_{\mathcal{O}} W_\sigma)^G$$

satisfies condition (3) of Definition 3.1. However, this is precisely the point of the representation W_σ: it cuts out the automorphic representations giving rise to minimal deformations of $\overline{\rho}$ at the primes in $T(\overline{\rho})$ (see [15, Lemma 5.1.1]). (Note that, since n is large, the space $H^0(X_{V_1(Q)}, \omega^n)_{\mathfrak{m}}$ is torsion free). This completes the proof. □

Theorem 1.4 follows from the previous result and Verdier duality as in Sect. 3.8. We remark that $H_0(X, \omega \otimes \mathcal{L}_\sigma)_{\mathfrak{m}_\emptyset}$ is of rank one over $\mathbf{T}_{\emptyset, \mathfrak{m}_\emptyset}$ when $H_0(X, \omega_K \otimes \mathcal{L}_\sigma)_{\mathfrak{m}_\emptyset}$ is non-zero. This follows from multiplicity one for GL(2) and [15, Lemma 4.2.4(3)].

4 Complements

4.1 Multiplicity two

Although this is not needed for our main results, we deduce in this section some facts about global multiplicity of Galois representations in modular Jacobians. Recall that k denotes a finite field of odd characteristic, $\mathcal{O}$ denotes the ring of integers of some finite extension K of $\mathbf{Q}_p$ with uniformizer ϖ and $\mathcal{O}/\varpi = k$.

We recall some standard facts about Cohen–Macaulay rings from [45], Sect. 21.3 (see also [46]). Let $(A, \mathfrak{m}, k)$ be a complete local Cohen–Macaulay ring of dimension n. Then A admits a canonical module ω_A. Moreover, if $(x_1, \ldots x_m)$ is a regular sequence for A, and $B = A/(x_1, \ldots, x_m)$, then

$$\omega_B := \omega_A \otimes_A B$$

is a canonical module for B. It follows that

$$\omega_A \otimes_A A/\mathfrak{m} = \omega_A \otimes_A (B \otimes_B B/\mathfrak{m}) = \omega_B \otimes_B B/\mathfrak{m}.$$

If $m = n$, so B is of dimension zero, then $\mathrm{Hom}(*, \omega_B)$ is a dualizing functor, and so

$$\dim_k B[\mathfrak{m}] = \dim_k \omega_B \otimes B/\mathfrak{m} = \dim_k \omega_A \otimes A/\mathfrak{m}.$$

Moreover, we have the following:

Lemma 4.1 *Let A be a finite flat local $\mathbf{Z}_p$-algebra. Suppose that A is Cohen–Macaulay. Then $\mathrm{Hom}_{\mathbf{Z}_p}(A, \mathbf{Z}_p)$ is a canonical module for A.*

Proof More generally, if A is a module-finite extension of a regular (or Gorenstein) local ring R, then (by Theorem 21.15 of [45]) $\mathrm{Hom}_R(A, R)$ is a canonical module for A. □

Finally, we note the following:

Lemma 4.2 *If B is a complete local Cohen–Macaulay $\mathcal{O}$-algebra and admits a dualizing module ω_B with μ generators, then the same is true for the power series ring $A = B[T_1, \ldots, T_n]$. Moreover, the same is also true for $B\widehat{\otimes}_{\mathcal{O}}C$, for any complete local $\mathcal{O}$-algebra C which is a complete intersection.*

Proof For power series rings this is a special case of the discussion above. Consider now the case of $B\widehat{\otimes}_{\mathcal{O}}C$. By assumption, C is a quotient of $\mathcal{O}[T_1, \ldots, T_n]$ by a regular sequence. Hence $B\widehat{\otimes}_{\mathcal{O}}C$ is a quotient of $B[T_1, \ldots, T_n]$ by a regular sequence, and the result follows from the discussion above applied to the maps $B[T_1, \ldots, T_n] \to B$ and $B[T_1, \ldots, T_n] \to B\widehat{\otimes}_{\mathcal{O}}C$ respectively. □

As an example, this applies to $B[\Delta]$ for any finite abelian group Δ of p-power order, since $\mathcal{O}[\Delta]$ is a complete intersection.

Let $\overline{\rho}: G_p \to \mathrm{GL}_2(k)$ be unramified with $\overline{\rho}(\mathrm{Frob}_p)$ scalar. Let $\widetilde{R}^{\dagger}$ denote the framed deformation ring of ordinary representations of weight n over $\mathcal{O}$-algebras with fixed determinant (together with a Frobenius eigenvalue α acting on an "unramified quotient") as in Theorem 3.19.

Theorem 4.3 *$\widetilde{R}^{\dagger}$ is a complete normal local Cohen–Macaulay ring of relative dimension 4 over $\mathcal{O}$. Let $\omega_{\widetilde{R}^{\dagger}}$ denote the canonical module of $\widetilde{R}^{\dagger}$. Then* $\dim_k \omega_{\widetilde{R}^{\dagger}}/\mathfrak{m} = 3$.

Proof Following the previous discussion, to determine $\dim_k \omega_{\widetilde{R}^{\dagger}}/\mathfrak{m}$, it suffices to find a regular sequence of length 5 ($= \dim \widetilde{R}^{\dagger}$), take the quotient C, and compute $\dim_k C[\mathfrak{m}]$. Since $\widetilde{R}^{\dagger}$ is $\mathcal{O}$-flat, ϖ is regular, and thus we may choose ϖ as the first term of our regular sequence. Yet the method of Snowden shows that $\widetilde{R}^{\dagger} \otimes k$ is given by the following relations (the completion of $\mathcal{B}_1$ at $b = (1, 1; 0)$ in the notation of [41]):

$$m = \begin{pmatrix} a & b \\ c & -a \end{pmatrix}, n = \phi - \mathrm{id} = \begin{pmatrix} \phi_1 & \phi_2 \\ \phi_3 & \phi_4 \end{pmatrix}, \beta = \alpha - 1,$$
$$mn = \beta m, \; P_{\phi}(\alpha) = 0, \; m^2 = 0, \; \det(\phi) = 1$$

Explicitly, in terms of equations, this is given by the quotient A of

$$k[a, b, c, \phi_1, \phi_2, \phi_3, \phi_4, \beta]$$

by the following relations:

$$\begin{aligned} &\phi_1 + \phi_4 + \phi_1\phi_4 - \phi_2\phi_3 = 0, \qquad \beta^2 - (\phi_1 + \phi_4)\beta - (\phi_1 + \phi_4) = 0, \\ &a\phi_1 + b\phi_3 = a\beta, a\phi_2 + b\phi_4 = b\beta, -a\phi_3 + c\phi_1 = c\beta, \\ &a\phi_4 - c\phi_2 = a\beta, a^2 + bc = 0. \end{aligned}$$

For a complete local k-algebra $(R, \mathfrak{m})$ with residue field k, let

$$H_R(t) = \sum_{n=0}^{\infty} \dim_k(\mathfrak{m}^n/\mathfrak{m}^{n+1})t^n \in \mathbf{Z}[t]$$

denote the corresponding Hilbert series. We define a partial ordering of elements of $\mathbf{Z}[t]$ as follows: say that

$$\sum_{n=0}^{\infty} a_n t^n \geq \sum_{n=0}^{\infty} b_n t^n$$

whenever $a_n \geq b_n$ for all n.

Lemma 4.4 *Let $x \in \mathfrak{m}^d$. We have*

$$\frac{H_{R/x}(t)}{1-t} \geq H_R(t) \cdot \frac{1-t^d}{1-t},$$

and equality holds if and only if x is a regular element. Moreover, if there is an isomorphism

$$R \simeq \mathrm{gr}(R) = \bigoplus \mathfrak{m}^n/\mathfrak{m}^{n+1},$$

and x is pure of degree d, then equality holds if and only if x is a regular element.

Proof There is an exact sequence as follows:

$$R/\mathfrak{m}^n \to R/\mathfrak{m}^n \to R/(x, \mathfrak{m}^n) \to 0.$$

The kernel of the first map certainly contains $\mathfrak{m}^{n-d}/\mathfrak{m}^n$. If $H_R(t) = \sum a_n t^n$ and $H_{R/x}(t) = \sum b_n t^n$, it follows that

$$\begin{aligned}
&\text{coefficient of } t^{m+d-1} \text{ in } \frac{H_{R/x}(t)}{(1-t)} \\
&= \sum_{n=0}^{m+d-1} b_n = \dim R/(x, \mathfrak{m}^{m+d}) \\
&= \dim \operatorname{coker}(R/\mathfrak{m}^{m+d} \to R/\mathfrak{m}^{m+d}) \\
&= \dim \ker(R/\mathfrak{m}^{m+d} \to R/\mathfrak{m}^{m+d}) \\
&\geq \dim \mathfrak{m}^m/\mathfrak{m}^{m+d} \\
&= a_m + a_{m+1} + \cdots + a_{m+d-1} \\
&= \text{coefficient of } t^{m+d-1} \text{ in } H_R(t)(1 + t + \cdots + t^{d-1}).
\end{aligned}$$

This proves the inequality. (Note that the coefficients of t^n for $n < d$ are automatically the same). On the other hand, assume that x is not a regular element. By assumption, there exists a non-zero element $y \in R$ such that $xy = 0$. By Krull's intersection theorem, there exists an m such that $y \notin \mathfrak{m}^m$. For such an m, it follows that the kernel of $R/\mathfrak{m}^{m+d} \to R/\mathfrak{m}^{m+d}$ is strictly bigger than $\mathfrak{m}^m/\mathfrak{m}^{m+d}$, and the inequality above is strict. Finally, assume that x is a regular element, and that $R \simeq \mathrm{gr}(R)$. Then the kernel of the map with $n = m + d$ above is precisely $\mathfrak{m}^m/\mathfrak{m}^{m+d}$, and we have equality.

Note that, in the non-graded case, the converse is not true, namely, x may be regular of degree one and yet the equality $H_{R/x}(t) = H_R(t)(1-t)$ fails; as an example one may take $R = k[\phi^6, \phi^7, \phi^{15}]$ and $x = \phi^6$. Then x is regular, but

$$\frac{H_{R/x}(t)}{1-t} = \frac{1+2t+2t^2+t^3}{1-t} > \frac{1+2t+t^2+t^3+t^5}{1-t} = H_R(t).$$

The point is that x is no longer regular in $\mathrm{gr}(R)$. (In fact, R is a Cohen–Macaulay domain, but the depth of $\mathrm{gr}(R)$ is zero; this example was taken from [47]).

Let $B \simeq A/\beta$. The equation $\beta^2 - (\phi_1+\phi_4)\beta - (\phi_1+\phi_4) = 0$ in A becomes $\phi_1 + \phi_4 = 0$ in B, and hence B is the quotient of

$$k[a, b, c, \phi_1, \phi_2, \phi_3]$$

by the following relations:

$$\begin{aligned} &-\phi_1^2 - \phi_2\phi_3 = 0,\\ &a\phi_1 + b\phi_3 = 0, a\phi_2 - b\phi_1 = 0, -a\phi_3 + c\phi_1 = 0, -a\phi_1\\ &\quad -c\phi_2 = 0, a^2 + bc = 0. \end{aligned}$$

All the relations in B are pure of degree two, and hence there is an isomorphism $B \simeq \mathrm{gr}(B)$.

Lemma 4.5 *The first few terms of $H_B(t)$ are*

$$H_B(t) = 1 + 6t + 15t^2 + \cdots$$

Proof Clearly $\dim(B/\mathfrak{m}) = 1$. B is a quotient of a power series ring $S = k[a, b, c, \phi_1, \phi_2, \phi_3]$. Moreover, since all the relations are quadratic, we have

$$\dim \mathfrak{m}/\mathfrak{m}^2 = \dim \mathfrak{m}_S/\mathfrak{m}_S^2 = 6.$$

The six generators of S give rise, *a priori*, to

$$\dim \mathfrak{m}_S^2/\mathfrak{m}_S^3 = \binom{7}{2} = 21$$

generators of $\mathfrak{m}^2/\mathfrak{m}^3$. Note, however, that we have 6 quadratic relations. In order to prove that $\dim \mathfrak{m}^2/\mathfrak{m}^3 = 21 - 6 = 15$, it suffices to show that these six relations are linearly independent. Choose a basis of $\mathfrak{m}_S^2/\mathfrak{m}_S^3$ coming from the lexiographic ordering $a > b > c > \phi_1 > \phi_2 > \phi_3$. With respect to this basis, the matrix of relations is as follows:

$$\begin{pmatrix} a^2 & 0 & 0 & 0 & 0 & 0 & 1 \\ ab & 0 & 0 & 0 & 0 & 0 & 0 \\ ac & 0 & 0 & 0 & 0 & 0 & 0 \\ a\phi_1 & 0 & 1 & 0 & 0 & -1 & 0 \\ a\phi_2 & 0 & 0 & 1 & 0 & 0 & 0 \\ a\phi_3 & 0 & 0 & 0 & -1 & 0 & 0 \\ b^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ bc & 0 & 0 & 0 & 0 & 0 & 1 \\ b\phi_1 & 0 & 0 & -1 & 0 & 0 & 0 \\ b\phi_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ b\phi_3 & 0 & 1 & 0 & 0 & 0 & 0 \\ c^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ c\phi_1 & 0 & 0 & 0 & 1 & 0 & 0 \\ c\phi_2 & 0 & 0 & 0 & 0 & -1 & 0 \\ c\phi_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ \phi_1^2 & -1 & 0 & 0 & 0 & 0 & 0 \\ \phi_1\phi_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \phi_1\phi_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ \phi_2^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \phi_2\phi_3 & -1 & 0 & 0 & 0 & 0 & 0 \\ \phi_3^2 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The minor consisting of rows 1, 4, 6, 9, 14, and 16 has determinant 1, and hence the result follows. □

Recall (Theorem 3.4.1 of [41]) that A is, in addition to being Cohen–Macaulay, also a domain. Since $\beta \neq 0$ (it is non-zero in $\mathfrak{m}/\mathfrak{m}^2$), it follows that β is a regular element, and hence $B = A/\beta$ is also Cohen–Macaulay.

Lemma 4.6 *If $I \subset B$ is an ideal generated by a regular sequence of elements of pure degree one of length 3, then*

$$H_{B/I}(t) = 1 + 3t.$$

Moreover, if I is any ideal generated by three pure elements of degree one such that $H_{B/I}(t) = 1 + 3t$, then the generators of I consist of a regular sequence.

Proof Let R be a complete local Cohen–Macaulay Noetherian graded k-algebra with residue field k. Replacing R by $R \otimes_k \overline{k}$ does not effect the Hilbert series of R. Assume that $\dim(R) \geq 1$, so that $\mathfrak{m}$ is not an associated prime. We claim that $R \otimes_k \overline{k}$ admits a regular element $x \in \mathfrak{m}$ of pure degree one. Without loss of generality, we assume that $k = \overline{k}$. The set of zero divisors is the union of the associated primes. By assumption, $\mathfrak{m}$ is not one of the associated primes. Hence, for every associated prime $\mathfrak{p}$, the image of $\mathfrak{p}$ in $\mathfrak{m}/\mathfrak{m}^2$ is proper (since otherwise $\mathfrak{p} = \mathfrak{m}$ by Nakayama's Lemma). Because R is Noetherian, there exist only finitely many associated primes. Hence the union of the images of all such $\mathfrak{p}$ cut out a finite number of proper linear subspaces of $\mathfrak{m}/\mathfrak{m}^2$. Since k is infinite, such a union misses an infinite number of points, and hence there exists an $x \in \mathfrak{m}\backslash\mathfrak{m}^2$ which is not a zero-divisor. By induction, there exists a regular sequence of length $\dim(R)$ generated by pure degree one elements. It follows that, after a finite extension, B admits a regular sequence of length 3 generated by pure degree one elements. By Lemma 4.4 (in the graded case), if I is the corresponding ideal, then

$$H_{B/I}(t) = H_B(t)(1-t)^3 = (1+6t+15t^2+\cdots)(1-t)^3 = 1+3t+O(t^3).$$

If $\mathfrak{m}$ is the maximal ideal of B/I, we deduce that $\mathfrak{m}^2/\mathfrak{m}^3 = 0$, and thus by Nakayama's Lemma that $\mathfrak{m}^2 = 0$, and $H_{B/I}(t) = 1+3t$. Conversely, if I is any ideal generated by three pure elements such that $H_{B/I}(t) = 1+3t$, then by Lemma 4.4, we deduce that the three generators of I consist of a regular sequence. □

Lemma 4.7 *$\{\beta, a, \phi_2+\phi_3, b+c+\phi_1\}$ is a regular sequence in A.*

Proof It suffices to show that $\{a, \phi_2+\phi_3, b+c+\phi_1\}$ is regular in $B = A/\beta$. By Lemma 4.6, it suffices to show that the Hilbert series of B/I with $I = (a, \phi_2+\phi_3, b+c+\phi_1)$ is $1+3t$. If $C = B/I$, then we compute that C is given by the quotient of

$$k[b, c, \phi_2]$$

by the following relations:

$$\begin{aligned} &-(b+c)^2+\phi_2^2 = 0,\\ &-b\phi_2 = 0,\ b(b+c) = 0,\ (b+c)c = 0,\ -c\phi_2 = 0,\ bc = 0. \end{aligned}$$

Let $x = b$, $y = c$, and $z = \phi_2$. Then, from the second, fifth, and sixth relations, we deduce that

$$xz = yz = xy = 0.$$

Combining this with the third and fourth equations yields:

$$x^2 = x^2+xy = 0,\ \ y^2 = xy+y^2 = 0.$$

The first equation yields

$$z^2 = -x^2-y^2-2xy+z^2 = 0.$$

It follows that C is a quotient of

$$k[x, y, z]/(x^2, y^2, z^2, xy, xz, yz).$$

On the other hand, since all the relations are trivial in $\mathfrak{m}^2$, we have $\dim(\mathfrak{m}/\mathfrak{m}^2) = 3$. Hence the Hilbert polynomial of C is $1+3t$, and the sequence is regular in A. □

Since $\dim C[\mathfrak{m}] = 3$, this completes the proof. □

From now until the end of Sect. 4.1, we let $\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(k)$ be an absolutely irreducible modular (= odd) representation of Serre conductor $N = N(\overline{\rho})$ and Serre weight $k(\overline{\rho})$ with $p+1 \geq k(\overline{\rho}) \geq 2$. This is an abuse of notation as we have already fixed a representation $\overline{\rho}$ in Sect. 3.1 but we hope it will not lead to confusion. Assume that $\overline{\rho}$ has minimal conductor amongst all its twists at all other primes (one can always twist $\overline{\rho}$ to satisfy these condition). One knows that $\overline{\rho}$ occurs as the mod-p reduction of a modular form of weight 2 and level N^*, where $N^* = N$ if $k = 2$ and Np otherwise. Let $\mathbf{T}$ denote the ring of endomorphisms of $J_1(N^*)/\mathbf{Q}$ generated by the Hecke operators T_l for *all* primes l (including p), and let $\mathfrak{m}$ denote the maximal ideal of $\mathbf{T}$ corresponding to $\overline{\rho}$. Assume that $p \geq 3$.

Theorem 4.8 *(Multiplicity one or two) If $\overline{\rho}$ is either ramified at p or unramified at p and $\overline{\rho}(\mathrm{Frob}_p)$ is non-scalar, then $J_1(N^*)[\mathfrak{m}] \simeq \overline{\rho}$, that is, $\mathfrak{m}$ has multiplicity one. If $\overline{\rho}$ is unramified at p and $\overline{\rho}(\mathrm{Frob}_p)$ is scalar, then $J_1(N^*)[\mathfrak{m}] \simeq \overline{\rho} \oplus \overline{\rho}$, that is, $\mathfrak{m}$ has multiplicity two.*

Remark 4.9 By results of Mazur [48] (Proposition 14.2), Mazur–Ribet [49] (Main Theorem), Gross [28] (Proposition 12.10), Edixhoven [31] (Thm. 9.2), Buzzard [50], and Wiese [51] Cor. 4.2, the theorem is known except in the case when $\overline{\rho}$ is unramified at p and $\overline{\rho}(\mathrm{Frob}_p)$ is scalar. In this case, Wiese [51] has shown that the multiplicity is always *at least* two. Thus our contribution to this result is to show that the multiplicity is *exactly* two in the scalar case.

Remark 4.10 It was historically the case that multiplicity one was an *ingredient* in modularity lifting theorems, e.g., Theorem 2.1 of [1]. It followed that the methods used to prove such theorems required a careful study of the geometry of $J_1(N^*)$. However, a refinement of the Taylor–Wiles method due to Diamond showed that one could *deduce* multiplicity one in certain circumstances while simultaneously proving a modularity theorem (see [26]). Our argument is in the spirit of Diamond, where it is the geometry of a local deformation ring rather than $J_1(N^*)$ that is the crux of the matter.

Proof Let G denote the part of the p-divisible group of $J_1(N^*)$ which is associated to $\mathfrak{m}$. By [28], Prop 12.9, as well as the proof of Prop 12.10, recall there is an exact sequence of groups

$$0 \to T_p G^0 \to T_p G \to T_p G^e \to 0$$

which is stable under $\mathbf{T}_{\mathfrak{m}}$. Moreover, $T_p G^0$ is free of rank one over $\mathbf{T}_{\mathfrak{m}}$, and $T_p G^e = \mathrm{Hom}(T_p G^0, \mathbf{Z}_p)$.

We may assume that $\overline{\rho}$ is unramified at p and $\overline{\rho}(\mathrm{Frob}_p)$ is scalar. Thus $N^* = pN$ (since p is odd). Let M denote the largest factor of N which is only divisible by the so called "harmless" primes, that is, the primes v such that $v \equiv 1 \mod p$ and such that $\overline{\rho}|G_v$ is absolutely irreducible. Define the group Φ as follows:

$$\Phi := \mathbf{Z}_p \otimes \prod_{x|M} (\mathbf{Z}/x\mathbf{Z})^\times$$

Φ measures the group of Dirichlet characters (equivalently, characters of $G_{\mathbf{Q}}$) congruent to $1 \mod \varpi$ which preserve the set of lifts of $\overline{\rho}$ of minimal conductor under twisting (by assumption, $\overline{\rho}$ has minimal conductor amongst its twists, so an easy exercise shows that these are the only twists with this property). Extending $\mathcal{O}$ if necessary, we may assume that each character $\phi \in \widehat{\Phi} := \mathrm{Hom}(\Phi, \overline{\mathbf{Q}}_p^\times)$ is valued in $\mathcal{O}^\times$. For $\phi \in \widehat{\Phi}$, and let χ_ϕ denote the character $\epsilon \cdot \langle \overline{\rho}\epsilon^{-1} \rangle \phi$ of $G_{\mathbf{Q}}$. For $v|N^*$ we define a quotient $R_v = R_{v,\phi}$ of the universal framed deformation ring with determinant χ_ϕ of $\overline{\rho}|G_v$ as follows:

(1) When $v = p$, $R_v = R_{v,\phi}$ is the ordinary framed deformation ring $R^\dagger$ of Sect. 3.7 (with $n = 2$).
(2) When $v \neq p$, $R_v = R_{v,\phi}$ is the unrestricted framed deformation ring with determinant $\chi_\phi|G_v$.

The isomorphism types of these deformation rings do not depend on ϕ. Let $R = R_\phi$ denote the (global) universal deformation ring of $\overline{\rho}$ corresponding to deformations with determinant χ_ϕ which are unramified outside N^* and which are classified (after a choice of framing) by R_v for each $v|N^*$. Let $R^\square$ denote the framed version of R, with framings at each place $v|N^*$. Let $\mathbf{T}_\phi^{\mathrm{an}}$ be the anaemic weight 2, level $\Gamma_1(N^*)$ ordinary Hecke algebra (so it does not contain U_p) which acts on

$$S_\phi := \bigoplus_{\chi} S_2^{\mathrm{ord}}(\Gamma_1(N^*), \chi, \mathcal{O}),$$

where χ runs over all the characters of $(\mathbf{Z}/Np\mathbf{Z})^\times$ with $\chi|\Phi = \phi$. Note that

$$S_2^{\mathrm{ord}}(\Gamma_1(N^*), \mathcal{O}) \otimes \mathbf{Q} = \bigoplus_{\widehat{\Phi}} S_\phi \otimes \mathbf{Q}.$$

The reason for dealing with the harmless primes in this manner is as follows. For all *non*-harmless $v \neq p$, the conductor-minimal deformation ring of $\overline{\rho}|G_v$ (which classifies deformations which appear at level $\Gamma_1(N)$) is isomorphic to the ring $R_{v,\phi}$ (up to unramified twists). Equivalently, for non-harmless primes, any characteristic zero lift ρ of $\overline{\rho}$ is uniquely twist equivalent to a lift of minimal conductor. The only time this is not true is when $v \equiv 1 \mod p$ and $\overline{\rho}$ has trivial invariants under I_v. Since we are assuming that $\overline{\rho}|_{G_v}$ is minimally ramified amongst all its twists, this only happens when $\overline{\rho}|_{G_v}$ is absolutely irreducible and $v \equiv 1 \mod p$. However, the naïve conductor-minimal deformation deformation ring at a harmless prime is equal to the unrestricted deformation ring and does not have fixed inertial determinant, and one needs the determinant to be fixed for the Taylor–Wiles method to work correctly.

Consider the Galois representation $\rho : G_{\mathbf{Q}} \rightarrow \mathrm{GL}_2(\mathbf{T}^{\mathrm{an}}_{\phi,\mathfrak{m}})$ associated to eigenforms in S_ϕ. The character $\epsilon^{-1} \det \rho$ can be regarded as a character $\chi : (\mathbf{Z}/Np\mathbf{Z})^\times \rightarrow (\mathbf{T}^{\mathrm{an}}_{\phi,\mathfrak{m}})^\times$ with $\chi|\Phi = \phi$. Let ψ denote the restriction of χ to $\mathbf{Z}_p \otimes \prod_{x \nmid M} (\mathbf{Z}/x\mathbf{Z})^\times$, which we may regard as a character of $G_{\mathbf{Q}}$. After twisting ρ by $\psi^{-1/2}$, we obtain a Galois representation

$$G_{\mathbf{Q}} \rightarrow \mathrm{GL}_2(\mathbf{T}^{\mathrm{an}}_{\phi,\mathfrak{m}})$$

with determinant $\chi_\phi = \epsilon \cdot \langle \overline{\rho} \epsilon^{-1} \rangle \phi$ which is classified by $R = R_\phi$. Our hypotheses on $\overline{\rho}$ (that $\overline{\rho}$ is absolutely irreducible and unramified at p) imply that $\overline{\rho}|G_{\mathbf{Q}(\zeta_p)}$ is absolutely irreducible. Kisin's improvement of the Taylor–Wiles method yields an isomorphism $R_\phi[1/p] \simeq \mathbf{T}^{\mathrm{an}}_{\phi,\mathfrak{m}}[1/p]$. (Here we apply the Taylor–Wiles type patching results Proposition 3.3.1 and Lemma 3.3.4 of [52]—as in the proof of Theorem 3.4.11 of *ibid.*—except that the rings denoted B and D in the statements of these results may no longer be integral domains in our situation (though their generic fibres will be formally smooth over K by Lemma 4.11 below). This is due to the fact that the rings $R_{v,\phi}$ defined above may have multiple irreducible components for certain $v \neq p$. On the other hand, the only place in [52] where the assumption that B and D be integral domains is used is in the first paragraph of the proof of Lemma 3.3.4. In our case, it will suffice to show that each irreducible component of $R_{v,\phi}$ is in the support of S_ϕ. This follows from now standard results on the existence of modular deformations with prescribed local inertial types). □

Now, $R^\square$ is (non-canonically) a power series ring over R, and is realized as a quotient of

$$\left(R^\dagger \widehat{\otimes} \widehat{\bigotimes}_{v|N} R_v\right)[x_1, \ldots, x_n] \rightarrow R^\square$$

by a sequence of elements that can be extended to a system of parameters; this last fact follows from the proof of Prop 5.1.1 of [41] and the fact that R is finite over $\mathcal{O}$ (for the most general results concerning the finiteness of deformation rings over $\mathcal{O}$, see Theorem 10.2 of [53]). As a variant of this, we may consider deformations of $\overline{\rho}$ together with an eigenvalue α of Frobenius at p. Globally, this now corresponds to a modified global deformation ring $\widetilde{R} = \widetilde{R}_{\phi}$ and the corresponding framed version $\widetilde{R}^{\square}$, where we now map to the full Hecke algebra $\mathbf{T}_{\mathfrak{m}}$. There are surjections:

$$R_{\mathrm{loc}}[x_1, \ldots, x_n] := \left(\widetilde{R}^{\dagger}\widehat{\otimes}\widehat{\bigotimes}_{v|N} R_v\right)[x_1, \ldots, x_n] \rightarrow \widetilde{R}^{\square} \rightarrow \widetilde{R} \rightarrow \widetilde{R}/\varpi.$$

Since $\widetilde{R}$ is finite over $\mathcal{O}$, it follows that $\widetilde{R}/\varpi$ is Artinian. Again, as in the proof of Prop 5.1.1 of [41] (see also Proposition 4.1.5 of [54]), the kernel of the composition of these maps is given by a system of parameters, one of which is ϖ. On the other hand, we have:

Lemma 4.11 *The rings $R_v = R_{v,\phi}$ for $v \neq p$ are complete intersections. Moreover, their generic fibres $R_v[1/p]$ are formally smooth over K.*

Proof There are three cases in which R_v is not smooth. In two of these cases, we shall prove that R_v is a power series ring over $\mathcal{O}[\Delta]$ for some finite cyclic abelian p-group Δ. Since $\mathcal{O}[\Delta]$ is manifestly a complete intersection with formally smooth generic fibre, this suffices to prove the lemma in these cases. In the other case, we will show that R_v is a quotient of a power series ring by a single relation. This shows that it is a complete intersection. The three situations in which R_v is not smooth correspond to primes v such that:

(1) $v \equiv 1 \mod p$, $\overline{\rho}|G_v$ is reducible, and $\overline{\rho}|I_v \simeq \chi \oplus 1$ for some ramified χ.
(2) $v \equiv -1 \mod p$, $\overline{\rho}|G_v$ is absolutely irreducible and induced from a character ξ.
(3) $v \equiv 1 \mod p$, $\overline{\rho}^{I_v}$ is 1-dimensional and $\overline{\rho}^{\mathrm{ss}}|G_v$ is unramified.

Suppose that v is a vexing prime (the second case). Any conductor-minimal deformation of $\overline{\rho}$ is induced from a character of the form $\langle\xi\rangle\psi$ over the quadratic unramified extension of $\mathbf{Q}_v$, where $\psi \mod \varpi$ is trivial. It follows that ψ is tamely ramified, and in particular, up to unramified twist, it may be identified with a character of $\mathbf{F}_{v^2}^{\times}$ of p-power order. We may therefore write down the universal deformation explicitly, which identifies R_v with a power series ring over $\mathcal{O}[\Delta]$, where Δ is the maximal p-quotient of $\mathbf{F}_{v^2}^{\times}$.

Suppose that we are in the first case, and so, after an unramified twist, $\overline{\rho}|G_v \cong \chi \oplus 1$. All R_v-deformations of $\overline{\rho}$ are of the form $(\langle\chi\rangle\psi \oplus \psi^{-1}) \otimes (\chi_{\phi}\langle\chi^{-1}\rangle)^{1/2}$, where $\psi \equiv 1 \mod \varpi$. It follows that ψ is tamely ramified, and in particular, decomposes as an unramified character and a character of $\mathbf{F}_v^{\times}$ of p-power order. We may therefore write down the universal framed deformation explicitly, which identifies R_v with a power series ring over $\mathcal{O}[\Delta]$, where Δ is the maximal p-quotient of $\mathbf{F}_v^\times$.

In the third case, the deformation rings are not quite as easy to describe explicitly, so we use a more general argument. As noted by the referee, the following argument may also be easily modified to deal with the first two cases. We first note that $R_{v,\phi}$ is a quotient of a power series ring over $\mathcal{O}$ in $\dim Z^1(G_v, \mathrm{ad}^0\overline{\rho}) = 4$ variables by at most $\dim H^2(G_v, \mathrm{ad}^0\overline{\rho}) = 1$ relation. Closed points on the generic fiber of $R_{v,\phi}$ correspond to lifts ρ of $\overline{\rho}|G_v$ which are either unramified twists of the Steinberg representation or lifts which decompose (after inverting p) into a sum $\chi_\phi\psi \oplus \psi^{-1}$ with $\psi|I_v$ of p-power order. The completion of $R_v[1/p]$ at such a point is the corresponding characteristic 0 deformation ring of the lift ρ (see Proposition 2.3.5 of [52]). In each case, we have $\dim H^2(G_v, \mathrm{ad}^0\rho) = 0$ and hence this ring is a power series ring (over the residue field at the point) in $\dim Z^1(G_v, \mathrm{ad}^0\rho) = 3$ variables. It follows that $R_v \cong \mathcal{O}[x_1, x_2, x_3, x_4]/(r)$ for some $r \neq 0$ and $R_v[1/p]$ is formally smooth over K[5]. This concludes the proof of the lemma. □

By Lemma 4.2, it follows that $R_{\mathrm{loc}}[x_1, \ldots, x_n]$ is Cohen–Macaulay, and hence the sequence of parameters giving rise to the quotient $\widetilde{R}/\varpi$ is a regular sequence. In particular, $\widetilde{R}$ is Cohen–Macaulay and ϖ-torsion free. Moreover, again by Lemma 4.2, the number of generators of the canonical module of of $R_{\mathrm{loc}}[x_1, \ldots, x_n]$ (and hence of $\widetilde{R}$) is equal to the number of generators of the canonical module of $\widetilde{R}^\dagger$, which is 3, by Theorem 4.3. Since patching arguments may also be applied to the adorned Hecke algebras $\mathbf{T}_\mathfrak{m}$, The method of Kisin yields an isomorphism $\widetilde{R}[1/p] = \mathbf{T}_{\phi,\mathfrak{m}}[1/p]$ (note that $\widetilde{R}^\dagger$ is a domain, and $\widetilde{R}^\dagger[1/p]$ is formally smooth). Since (as proven above) $\widetilde{R}$ is $\mathcal{O}$-flat, it follows that $\widetilde{R} \simeq \mathbf{T}_{\phi,\mathfrak{m}}$. In particular, we deduce that $\mathbf{T}_{\phi,\mathfrak{m}}$ is Cohen–Macaulay, and that $\dim \omega_{\mathbf{T}_{\phi,\mathfrak{m}}}/\mathfrak{m} = 3$. There is an isomorphism as follows:

$$S_2(\Gamma_1(N^*), \mathcal{O})_\mathfrak{m} \otimes K = \bigoplus_{\widehat{\Phi}} S_{\phi^2,\mathfrak{m}} \otimes K,$$

where, since p is odd, we write every element of $\widehat{\Phi}$ uniquely as a square. If $\mathbf{T}_{\phi,\mathfrak{m}}$ denotes the Hecke action on $S_{\phi,\mathfrak{m}} \otimes K$, then twisting by ϕ induces an isomorphism $\mathbf{T}_{\phi^2,\mathfrak{m}} \simeq \mathbf{T}_{1,\mathfrak{m}} \otimes_\mathcal{O} \mathcal{O}(\phi)$, since this is precisely the effect twisting has on the action of the diamond operators. (Here $\mathcal{O}(\phi) = \mathcal{O}$ with

[5] One may take r to be $C(T) - T$, where C is the Chebyshev-type polynomial determined by the relation $C(t + t^{-1}) = t^v + t^{-v}$, and T is the trace of a generator of tame inertia (note that $T - 2 \in \mathfrak{m}_{R_v}$). The generic fibre of R_v has $(q+1)/2$ geometric components, where q is the largest power of p dividing $v - 1$ (see also Theorem 1.0.1(A2.2) of [55]). One component corresponds to lifts of $\overline{\rho}$ on which inertia is nilpotent, and in particular has trace $T = 2$. The remaining $(q-1)/2$ components correspond to representations which are finitely ramified of order dividing q, on which $T = \zeta + \zeta^{-1}$ for some primitive q-th root of unity $\zeta \neq 1$.

the Φ action twisted by ϕ). If $\mathbf{T}_{\mathfrak{m}}$ is the Hecke ring at full level $\Gamma_1(N^*)$, then the restriction map induces an inclusion map

$$\mathbf{T}_{\mathfrak{m}} \hookrightarrow \bigoplus \mathbf{T}_{\phi^2,\mathfrak{m}} \simeq \mathbf{T}_{1,\mathfrak{m}} \otimes \bigoplus \mathcal{O}(\phi).$$

Since $\mathbf{T}_{\mathfrak{m}}$ is local, the image lands inside $\mathbf{T}_{1,\mathfrak{m}} \otimes \mathcal{O}[\Phi]$. Because the map above is an isomorphism after tensoring with K, and since all the relevant spaces of modular forms are ϖ-torsion free, we have an isomorphism $\mathbf{T}_{\mathfrak{m}} \simeq \mathbf{T}_{1,\mathfrak{m}} \otimes_{\mathcal{O}} \mathcal{O}[\Phi]$. Hence, applying Lemma 4.2 once more, we deduce that $\mathbf{T}_{\mathfrak{m}}$ is Cohen–Macaulay and $\dim \omega_{\mathbf{T}_{\mathfrak{m}}}/\mathfrak{m} = 3$

Since $\mathbf{T}_{\mathfrak{m}}$ is finite over $\mathbf{Z}_p$, we deduce by Lemma 4.1 that $\mathrm{Hom}(\mathbf{T}_{\mathfrak{m}}, \mathbf{Z}_p)$ is the canonical module of $\mathbf{T}_{\mathfrak{m}}$, and thus $\mathrm{Hom}(\mathbf{T}_{\mathfrak{m}}, \mathbf{Z}_p)/\mathfrak{m}$ also has dimension three. Yet we have identified $\mathrm{Hom}(\mathbf{T}_{\mathfrak{m}}, \mathbf{Z}_p)$ with $T_p G^e$, and it follows that $\dim G^0[\mathfrak{m}] = \dim T_p G^e/\mathfrak{m} = 3$, and hence

$$\dim J_1(N^*)[\mathfrak{m}] = \frac{1}{2}\dim G[\mathfrak{m}] = \frac{1}{2}(\dim G^0[\mathfrak{m}] + \dim G^e[\mathfrak{m}]) = \frac{3+1}{2} = 2.$$

□

Remark 4.12 If $X_H(N^*) = X_1(N^*)/H$ is the smallest quotient of $X_1(N^*)$ where one might expect $\overline{\rho}$ to occur, a similar argument shows that $J_H(N^*)[\mathfrak{m}]$ has multiplicity two if $\overline{\rho}$ is unramified and scalar at p, and has multiplicity one otherwise, providing that $p \neq 3$ and $\overline{\rho}$ is not induced from a character of $\mathbf{Q}(\sqrt{-3})$. The only extra ingredient required is the result of Carayol (see [56], Proposition 3 and also [57], Proposition 1.10).

Remark 4.13 We expect that these arguments should also apply in principle when $p = 2$; the key point is that one should instead use the quotient $\widetilde{R}^{\dagger}_3$ of $\widetilde{R}^{\dagger}$ (in the notation of [41], Sect. 4), corresponding to *crystalline* ordinary deformations. The special fibre of $\widetilde{R}^{\dagger}_3$ is (in this case) also given by $\mathcal{B}_1$, and thus one would deduce that the multiplicity of $\overline{\rho}$ is two when $\overline{\rho}(\mathrm{Frob}_2)$ is scalar, assuming that $\overline{\rho}$ is not induced from a quadratic extension. The key point to check is that the arguments above are compatible with the modifications to the $R = \mathbf{T}$ method for $p = 2$ developed by Khare–Wintenberger and Kisin (in particular, this will require that $\overline{\rho}$ is not dihedral).

4.2 Finiteness of deformation rings

Lemma 4.14 *Let $F/\mathbf{Q}$ be a number field, let k be a finite field, and let S denote a finite set of places not containing any $v|p$. Let $G_{F,S}$ denote the Galois group of the maximal extension of F unramified outside S. Let*

$$\overline{\rho} : G_{F,S} \to \mathrm{GL}_2(k)$$

be a continuous absolutely irreducible representation, and let R denote the universal deformation ring of $\overline{\rho}$. Suppose that the Galois representation associated to any $\overline{\mathbf{Q}}_p$-point of R has finite image, and suppose that there are only finitely many $\overline{\mathbf{Q}}_p$-points of R. Then $R[1/p]$ is reduced; equivalently, $R[1/p]^{\mathrm{red}} = R[1/p]$.

Proof Because S is a finite set of primes, it follows from the discussion in Sect. 1 p. 387 of [58] that R is a complete local Noetherian $W(k)$-algebra. The assumption that R has only finitely many $\overline{\mathbf{Q}}_p$-points implies that $R[1/p]^{\mathrm{red}}$ is isomorphic to a product of finitely many fields indexed by prime ideals $\mathfrak{p}^{\mathrm{red}}$ of $R[1/p]^{\mathrm{red}}$. Since $\mathrm{Spec}(R[1/p]^{\mathrm{red}})$ and $\mathrm{Spec}(R[1/p])$ are naturally isomorphic as sets, there is a bijection between primes $\mathfrak{p}$ of $R[1/p]$ and $\mathfrak{p}^{\mathrm{red}}$ of $R[1/p]^{\mathrm{red}}$. Hence $R[1/p]$ is a Noetherian semi-local ring, which therefore decomposes as a direct sum of its localizations over all finite ideals $\mathfrak{p}$. It suffices to show that the localizations of $R[1/p]$ and $R[1/p]^{\mathrm{red}}$ at every prime $\mathfrak{p}$ are isomorphic. Denote this localization of $R[1/p]$ by $(A, \mathfrak{m})$. Note that $A/\mathfrak{m} = A/\mathfrak{p} = R[1/p]^{\mathrm{red}}/\mathfrak{p}^{\mathrm{red}} \simeq E$ for some finite extension E of $\mathbf{Q}_p$. We have Galois representations as follows:

$$G_{F,S} \to \mathrm{GL}_2(R) \to \mathrm{GL}_2(R[1/p]) \to \mathrm{GL}_2(A) \to \mathrm{GL}_2(A/\mathfrak{m}^2) \to \mathrm{GL}_2(E)$$

To show that $A = E$, it suffices, by Nakayama's Lemma, to show that $A/\mathfrak{m}^2 = A/\mathfrak{m}$. Because E is of characteristic zero, the map $A/\mathfrak{m}^2 \to E$ splits, and $A/\mathfrak{m}^2$ has the structure of an E-algebra. If $A/\mathfrak{m}^2 \neq A/\mathfrak{m}$, then the map $A/\mathfrak{m}^2 \to A/\mathfrak{m}$ factors through a surjection $A/\mathfrak{m}^2 \to E[\epsilon]/\epsilon^2$. Because $\overline{\rho}$ is absolutely irreducible, the ring R is generated by the traces of the images of elements of $G_{F,S}$ (Proposition 4, S1.8 of [58]). It follows that the traces of the elements of $G_{F,S}$ generate $R[1/p]$ and all its quotients over $W(k) \otimes \mathbf{Q}$. It thus suffices to show that the images of the elements of $G_{F,S}$ in $\mathrm{GL}_2(E[\epsilon]/\epsilon^2)$ all have traces in E. Consider the corresponding Galois representation

$$\rho : G_{F,S} \to \mathrm{GL}_2(E[\epsilon]/\epsilon^2).$$

The composite to $\mathrm{GL}_2(E)$ has finite image by assumption. Denote the corresponding finite image Galois representation over E by V. Hence ρ arises from some extension

$$0 \to V \to W \to V \to 0.$$

Consider the restriction of this representation to a finite extension L/F such that $G_{L,S}$ acts trivially on V. Then the action of G_L on W factors through a $\mathbf{Z}_p$-extension which is unramified outside primes outside those above S, and is in particular unramified at all primes $v|p$. Such extensions are trivial by class field theory. Hence the extension splits over $G_{L,S}$. However, because $G_{L,S}$ has finite image in $G_{F,S}$, the extension also splits over $G_{F,S}$, because the inflation map is injective (the kernel is computed by H^1 of a finite group acting in characteristic zero). It follows that the extension is trivial over $G_{F,S}$. Yet this implies that the trace of the image of any element lies in E, which completes the proof. □

If $\overline{\rho} : G_{Q,S} \to \mathrm{GL}_2(k)$ as above is modular, then one can often deduce the assumptions (and hence the conclusions) of Lemma 4.14 from work of Buzzard–Taylor and Buzzard [22,23].

5 Imaginary quadratic fields

In this section, we apply our methods to Galois representations of regular weight over imaginary quadratic fields. The argument, formally, is very similar to what happens to weight one Galois representations over $G_{\mathbf{Q}}$. The most important difference is that we are not able to prove the existence of Galois representations associated to torsion classes in cohomology, and so our results are predicated on a conjecture that suitable Galois representations exist (Conjecture A).

5.1 Deformations of Galois representations

Let F be an imaginary quadratic field, and let $p \geq 3$ be a prime that is unramified in F. Suppose that $v|p$ is a place of F and A is an Artinian local $\mathcal{O}$-algebra. We say that a continuous representation $\rho : G_v \to \mathrm{GL}_2(A)$ is *finite flat* if there is a finite flat group scheme $\mathcal{F}/\mathcal{O}_{F_v}$ such that $\rho \cong \mathcal{F}(\overline{F}_v)$ as $\mathbf{Z}_p[G_v]$-modules, and $\det(\rho|I_v)$ is the cyclotomic character. We say that ρ is *ordinary* if ρ is conjugate in $\mathrm{GL}_2(A)$ to a representation of the form

$$\begin{pmatrix} \epsilon\chi_1 & * \\ 0 & \chi_2 \end{pmatrix}$$

where χ_1 and χ_2 are unramified.

Let

$$\overline{\rho} : G_F \to \mathrm{GL}_2(k)$$

be a continuous Galois representation such that the restriction

$$\overline{\rho} : G_{F(\zeta_p)} \to \mathrm{GL}_2(k)$$

is absolutely irreducible. Let $S(\overline{\rho})$ denote the set of primes not dividing p where $\overline{\rho}$ is ramified. We assume the following:

(1) $\det(\overline{\rho})$ is the mod-p reduction of the cyclotomic character.
(2) $\overline{\rho}$ is either ordinary or finite flat at $v|p$.
(3) If $x \in S(\overline{\rho})$, then either:
  (a) $\overline{\rho}|I_x$ is irreducible.
  (b) $\overline{\rho}|I_x$ is unipotent.
  (c) $\overline{\rho}|D_x$ is reducible, and $\rho|I_x$ is of the form $\psi \oplus \psi^{-1}$.
  (d) If $\overline{\rho}|D_x$ is irreducible and $\overline{\rho}|I_x$ is reducible, then $N_{F/\mathbf{Q}}(x) \not\equiv -1 \mod p$.

Let Q denote a finite set of primes in $\mathcal{O}_F$ not containing any primes above p and not containing any primes at which $\overline{\rho}$ is ramified. For objects R in $\mathcal{C}_\mathcal{O}$, we say that a representation $\rho : G \to \mathrm{GL}_2(R)$ is unipotent if, after some change of basis, the image of ρ is a subgroup of the matrices of the form $\begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}$. For objects R in $\mathcal{C}_\mathcal{O}$, we may consider lifts $\rho : G_F \to \mathrm{GL}_2(R)$ of $\overline{\rho}$ with the following properties:

(1) $\det(\rho) = \epsilon$.
(2) If $v|p$, then $(\rho \otimes_R (R/\mathfrak{m}_R^n))|G_v$ is finite flat or ordinary for all $n \geq 1$.
(3) If $v|p$ and $\overline{\rho}|G_v$ is finite flat, then $(\rho \otimes_R (R/\mathfrak{m}_R^n))|G_v$ is finite flat for all $n \geq 1$.
(4) If $x \notin Q \cup S(\overline{\rho}) \cup \{v|p\}$, then $\rho|G_x$ is unramified.
(5) **S** : If $x \in S(\overline{\rho})$, and $\overline{\rho}|I_x$ is unipotent, then $\rho|I_x$ is unipotent.
(6) **P** : If $x \in S(\overline{\rho})$, and $\overline{\rho}|I_x \simeq \psi \oplus \psi^{-1}$, then $\rho|I_x \simeq \langle\psi\rangle \oplus \langle\psi\rangle^{-1}$.
(7) **M** : If $x \in S(\overline{\rho})$, $\overline{\rho}|D_x$ is irreducible, and $\overline{\rho}|I_x = \psi_1 \oplus \psi_2$ is reducible, then $\rho|I_x = \langle\psi_1\rangle \oplus \langle\psi_2\rangle$.
(8) **H**: If $\overline{\rho}|I_x$ is irreducible, then $\rho(I_x) \xrightarrow{\sim} \overline{\rho}(I_x)$. (This also follows automatically from the determinant condition).

In cases 6 and 7 (and 8), there is an isomorphism $\rho(I_x) \xrightarrow{\sim} \overline{\rho}(I_x)$. For $x \in S(\overline{\rho})$, we say that $\overline{\rho}|D_x$ is of type **S**pecial, **P**rincipal, **M**ixed, or **H**armless respectively if is of the type indicated above. Note that primes of type **M** are called vexing by [14], but we have eliminated the most troublesome of the vexing primes, namely those x with $N_{F/\mathbf{Q}}(x) \equiv -1 \mod p$. The corresponding deformation functor is represented by a complete Noetherian local $\mathcal{O}$-algebra R_Q (this follows from the proof of Theorem 2.41 of [27]). If $Q = \emptyset$, we will sometimes denote R_Q by $R^{\min}$. Let $H^1_Q(F, \mathrm{ad}^0\overline{\rho})$ denote the Selmer group defined as the kernel of the map

$$H^1(F, \mathrm{ad}^0\overline{\rho}) \longrightarrow \bigoplus_x H^1(F_x, \mathrm{ad}^0\overline{\rho})/L_{Q,x}$$

where x runs over all primes of F and

- $L_{Q,x} = H^1(G_x/I_x, (\mathrm{ad}^0\overline{\rho})^{I_x})$ if $x \notin Q \cup \{v|p\}$;
- $L_{Q,x} = H^1(F_x, \mathrm{ad}^0\overline{\rho})$ if $x \in Q$ and $x \nmid p$;
- $L_{Q,v} = H^1_{\mathrm{f}}(F_v, \mathrm{ad}^0\overline{\rho})$ if $v|p$ and $v \notin Q$;

(The group $H^1_{\mathrm{f}}(F_v, \mathrm{ad}^0\overline{\rho})$ is defined as in Sect. 2.4 of [27]). Let $H^1_Q(F, \mathrm{ad}^0 \overline{\rho}(1))$ denote the corresponding dual Selmer group.

Proposition 5.1 *The reduced tangent space* $\mathrm{Hom}(R_Q/\mathfrak{m}_{\mathcal{O}}, k[\epsilon]/\epsilon^2)$ *of* R_Q *has dimension at most*

$$\dim_k H^1_Q(F, \mathrm{ad}^0\overline{\rho}(1)) - 1 + \sum_{x\in Q} \dim_k H^0(F_x, \mathrm{ad}^0\overline{\rho}(1)).$$

Proof The argument follows along the exact lines of Corollary 2.43 of [27]. The only difference in the calculation occurs at $v|p$ and at $v = \infty$. Specifically, when $v|p$ and p splits, the contribution to the Euler characteristic formula (Theorem 2.19 of [27]) is

$$\sum_{v|p} (\dim_k H^1_{\mathrm{f}}(F_v, \mathrm{ad}^0\overline{\rho}) - \dim_k H^0(F_v, \mathrm{ad}^0\overline{\rho})),$$

which, by Proposition 2.27 of [27], is at most 2. However, the contribution at the prime at ∞ is $-\dim_k H^0(\mathbf{C}, \mathrm{ad}^0\overline{\rho}) = -3$. When p is inert, the contribution at p is

$$\dim_k H^1_{\mathrm{f}}(F_p, \mathrm{ad}^0\overline{\rho}) - \dim_k H^0(F_p, \mathrm{ad}^0\overline{\rho})$$

which is also at most 2 (see, for instance, Corollary 2.4.3 of [3] and note that there is an inclusion $H^1(G_{F_p}/I_{F_p}, k) \subset H^1_{\mathrm{f}}(F_p, \mathrm{ad}\overline{\rho}) \cap H^1(F_p, k)$ where we view k as the scalar matrices in $\mathrm{ad}\overline{\rho}$). □

Suppose that $N_{F/\mathbf{Q}}(x) \equiv 1 \mod p$ and $\overline{\rho}(\mathrm{Frob}_x)$ has distinct eigenvalues for each $x \in Q$. Then $H^0(F_x, \mathrm{ad}^0\overline{\rho})$ is one dimensional for $x \in Q$ and the preceding proposition shows that the reduced tangent space of R_Q has dimension at most

$$\dim_k H^1_Q(F, \mathrm{ad}^0\overline{\rho}(1)) - 1 + \#Q.$$

We now show that one may choose a judicious set of primes (colloquially referred to as Taylor–Wiles primes) to annihilate the dual Selmer group.

Proposition 5.2 *Let* $q = \dim_k H^1_{\emptyset}(F, \mathrm{ad}^0\overline{\rho}(1))$ *and suppose that* $\overline{\rho}|G_{F(\zeta_p)}$ *is absolutely irreducible. Then* $q \geq 1$ *and for any integer* $N \geq 1$ *we can find a set* Q_N *of primes of* F *such that*

(1) $\#Q_N = q$.
(2) $N_{F/\mathbf{Q}}(x) \equiv 1 \mod p^N$ *for each* $x \in Q_N$.
(3) *For each* $x \in Q_N$, $\overline{\rho}$ *is unramified at* x *and* $\overline{\rho}(\mathrm{Frob}_x)$ *has distinct eigenvalues.*
(4) $H^1_{Q_N}(F, \mathrm{ad}^0\overline{\rho}(1)) = (0)$.

In particular, the reduced tangent space of R_{Q_N} *has dimension at most* $q - 1$ *and* R_{Q_N} *is a quotient of a power series ring over* $\mathcal{O}$ *in* $q - 1$ *variables.*

Proof That $q \geq 1$ follows immediately from Proposition 5.1. Now suppose that Q is a finite set of primes of F containing no primes dividing p and no primes where $\overline{\rho}$ is ramified. Suppose that $\overline{\rho}(\mathrm{Frob}_x)$ has distinct eigenvalues and $N_{F/\mathbf{Q}}(x) \equiv 1 \mod p$ for each $x \in Q$. Then we have an exact sequence

$$0 \longrightarrow H^1_Q(F, \mathrm{ad}^0\overline{\rho}(1)) \longrightarrow H^1_\emptyset(F, \mathrm{ad}^0\overline{\rho}(1)) \longrightarrow \bigoplus_{x \in Q} H^1(G_x/I_x, \mathrm{ad}^0\overline{\rho}(1)).$$

Moreover, for each $x \in Q$, the space $H^1(G_x/I_x, \mathrm{ad}^0\overline{\rho}(1))$ is one-dimensional over k and is isomorphic to $\mathrm{ad}^0\overline{\rho}/(\overline{\rho}(\mathrm{Frob}_x) - 1)(\mathrm{ad}^0\overline{\rho})$ via the map which sends a class $[\gamma]$ to $\gamma(\mathrm{Frob}_x)$. It follows that we may ignore condition (1): if we can find a set $\tilde{Q}_N$ satisfying conditions (2), (3) and (4), then $\#\tilde{Q}_N \geq q$ and by removing elements of $\tilde{Q}_N$ if necessary, we can obtain a set Q_N satisfying (1)–(4).

By the Chebotarev density theorem, it therefore suffices to show that for each non-zero class $[\gamma] \in H^1_\emptyset(F, \mathrm{ad}^0\overline{\rho}(1))$, we can find an element $\sigma \in G_F$ such that

- $\sigma | G_{F(\zeta_{p^N})} = 1$;
- $\overline{\rho}(\sigma)$ has distinct eigenvalues;
- $\gamma(\sigma) \notin (\overline{\rho}(\sigma) - 1)(\mathrm{ad}^0\overline{\rho})$.

The existence of such a σ can be established exactly as in the proof of Theorem 2.49 of [27]. □

5.2 Homology of arithmetic quotients

Let $\mathbf{A}$ denote the adeles of $\mathbf{Q}$, and $\mathbf{A}^\infty$ the finite adeles. Similarly, let $\mathbf{A}_F$ and $\mathbf{A}_F^\infty$ denote the adeles and finite adeles of F. Let $\mathbf{G} = \mathrm{Res}_{F/\mathbf{Q}}\mathrm{PGL}(2)$, and write $G_\infty = \mathbf{G}(\mathbf{R}) = \mathrm{PGL}_2(\mathbf{C})$. Let K_∞ denote a maximal compact of G_∞ with connected component K_∞^0. For any compact open subgroup K of $\mathbf{G}(\mathbf{A}^\infty)$, we may define an arithmetic orbifold $Y(K)$ as follows:

$$Y(K) := \mathbf{G}(\mathbf{Q})\backslash\mathbf{G}(\mathbf{A})/K_\infty^0 K.$$

Remark 5.3 If K is a sufficiently small (neat) compact subgroup, then $Y(K)$ is a manifold. Moreover, it will also be a a (disjoint union of) $K(\pi, 1)$ spaces, since each component is the quotient of a contractible space. Recall that for a $K(\pi, 1)$-manifold M, there is a functorial isomorphism

$$H^n(\pi_1(M), \star) \simeq H^n(M, \star)$$

for all n. For orbifolds $M = \Gamma\backslash\mathbf{H}$ of a similar shape (with contractible $\mathbf{H}$), the cohomology of M *as an orbifold* satisfies the same formula. Note that the cohomology in this sense may differ from the cohomology of the underlying space. (For example, the underlying manifold of $\mathrm{PSL}_2(\mathbf{Z})$ is the punctured sphere which is contractible, whereas the underlying orbifold has interesting cohomology). We take the convention that, for any K, the cohomology of $Y(K)$ is understood to be the cohomology in the orbifold sense, namely, that the cohomology of each component is the cohomology of the corresponding arithmetic lattice. The main advantage of this approach is that, for any finite index normal subgroup $K' \trianglelefteq K$, the corresponding map of orbifolds

$$Y(K') \to Y(K)$$

is a covering map with Galois group K/K'. This approach is the analogue (in the world of PEL Shimura varieties) of working with stacks rather than the underlying schemes at non-representable level.

We will specifically be interested in the following K. Let $S(\overline{\rho})$ and Q be as above.

5.2.1 Arithmetic quotients

If v is a place of F and $c \geq 1$ is an integer, we define

$$\Gamma_0(v^c) = \left\{ g \in \mathrm{PGL}_2(\mathcal{O}_v) \mid g \equiv \begin{pmatrix} 1 & * \\ 0 & * \end{pmatrix} \mod \pi_v^c \right\}$$
$$\Gamma_1(v^c) = \left\{ g \in \mathrm{PGL}_2(\mathcal{O}_v) \mid g \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \mod \pi_v^c \right\}$$
$$\Gamma_p(v^c) = \left\{ g \in \mathrm{PGL}_2(\mathcal{O}_v) \mid g \equiv \begin{pmatrix} 1 & * \\ 0 & d \end{pmatrix} \mod \pi_v^c, d \text{ has } p- \text{ power order} \right\}$$

Let $K_Q = \prod_v K_{Q,v}$ and $L_Q = \prod_v L_{Q,v}$ denote the open compact subgroups of $\mathbf{G}(\mathbf{A})$ such that:

(1) If $v \in Q$, $K_{Q,v} = \Gamma_1(v)$.
(2) If $v \in Q$, $L_{Q,v} = \Gamma_0(v)$.
(3) If v is not in $S(\overline{\rho}) \cup \{v|p\} \cup Q$, then $K_{Q,v} = L_{Q,v} = \mathrm{PGL}_2(\mathcal{O}_v)$.
(4) If $v|p$, then $K_{Q,v} = L_{Q,v} = \mathrm{GL}_2(\mathcal{O}_v)$ if $\overline{\rho}|D_v$ is finite flat. Otherwise, $K_{Q,v} = L_{Q,v} = \Gamma_0(v)$.
(5) If $v \in S(\overline{\rho})$, $K_{Q,v} = L_{Q,v}$ is defined as follows:
   (a) If $\overline{\rho}$ is of type **S** at v, then $K_{Q,v} = \Gamma_0(v)$.
   (b) If $\overline{\rho}$ is of type **P**, **M** or **H** at v, then $K_{Q,v} = \Gamma_p(v^c)$, where c is the conductor of $\overline{\rho}|D_v$.

We define the arithmetic quotients $Y_0(Q)$ and $Y_1(Q)$ to be $Y(L_Q)$ and $Y(K_Q)$ respectively. These spaces are the analogues of the modular curves corresponding to the congruence subgroups consisting of $\Gamma_0(Q)$ and $\Gamma_1(Q)$ intersected with a level specifically tailored to the ramification structure of $\overline{\rho}$. Topologically, they are a finite disconnected union of finite volume arithmetic hyperbolic 3-orbifolds.

5.2.2 Hecke operators

We recall the construction of the Hecke operators. Let $g \in \mathbf{G}(\mathbf{A}^\infty)$ be an invertible matrix. For $K \subset \mathbf{G}(\mathbf{A}^\infty)$ a compact open subgroup, the Hecke operator $T(g)$ is defined on the homology modules $H_\bullet(Y(K), \mathcal{O})$ by considering the composition:

$$H_\bullet(Y(K), \mathcal{O}) \to H_\bullet(Y(gKg^{-1} \cap K), \mathcal{O}) \to H_\bullet(Y(K \cap g^{-1}Kg), \mathcal{O}) \\ \to H_\bullet(Y(K), \mathcal{O}),$$

the first map coming from the corestriction (= transfer) map, the second coming from the map $Y(gKg^{-1} \cap K, \mathcal{O}) \to Y(K \cap g^{-1}Kg, \mathcal{O})$ induced by right multiplication by g on $\mathbf{G}(\mathbf{A})$ and the third coming from the natural map on homology. (We recall that, since we are viewing these spaces as orbifolds, the map $Y(gKg^{-1} \cap K) \to Y(K)$ is always a covering map). The Hecke operators act on $H_\bullet(Y(K), \mathcal{O})$ but do not preserve the homology of the connected components. The group of components is isomorphic, via the determinant map, to

$$F^\times \backslash \mathbf{A}_F^{\infty,\times} / \mathbf{A}_F^{\infty,\times 2} \det(K).$$

This is the mod-2 reduction of a ray class group. For $\alpha \in \mathbf{A}_F^{\infty,\times}$, we define the Hecke operator T_α by taking

$$g = \begin{pmatrix} \alpha & 0 \\ 0 & 1 \end{pmatrix}.$$

If $\alpha \in \mathbf{A}_F^{\infty,\times}$ is a unit at all finite places, we denote the corresponding operator by $\langle \alpha \rangle$ and refer to it as a diamond operator; it acts as an automorphism on $Y(K)$ for all the K considered above.

Definition 5.4 Let $\mathbf{T}_Q^{\mathrm{an}}$ denote the sub-$\mathcal{O}$-algebra of $\mathrm{End}_{\mathcal{O}} H_1(Y_1(Q), \mathcal{O})$ generated by Hecke endomorphisms T_α for all α which are trivial at primes in $Q \cup S(\overline{\rho}) \cup \{v|p\}$. Let $\mathbf{T}_Q$ denote the $\mathcal{O}$-algebra generated by the same operators together with T_α for α non-trivial at places in Q. If $Q = \emptyset$, we write $\mathbf{T} = \mathbf{T}_\emptyset$ for $\mathbf{T}_Q$.

These rings are commutative. If $\epsilon \in \mathcal{O}_F^\times$ is a global unit, then T_ϵ acts by the identity. If $\mathfrak{a} \subseteq \mathcal{O}_F$ is an ideal prime to the level, we may define the Hecke operator $T_\mathfrak{a}$ as T_α where $\alpha \in \mathbf{A}_F^{\times,\infty}$ is any element which represents the ideal $\mathfrak{a}$ and such that α is 1 for each component dividing the level. In particular, if $\mathfrak{a} = x$ is prime, then T_x is uniquely defined when x is prime to the level but not when x divides the level.

5.3 Conjectures on existence of Galois representations

Let $\mathfrak{m}$ denote a maximal ideal of $\mathbf{T}_Q$, and let $\mathbf{T}_{Q,\mathfrak{m}}$ denote the completion. It is a local ring which is finite (but not necessarily flat) over $\mathcal{O}$.

Definition 5.5 We say that $\mathfrak{m}$ is *Eisenstein* if $T_\lambda - 2 \in \mathfrak{m}$ for all but finitely primes λ which split completely in some fixed abelian extension of F. We say that $\mathfrak{m}$ is *non-Eisenstein* if it is not Eisenstein.

We say that $\mathfrak{m}$ is *associated to* $\overline{\rho}$ if for each $\lambda \notin S(\overline{\rho}) \cup Q \cup \{v|p\}$, we have an inclusion $T_\lambda - \mathrm{Trace}(\overline{\rho}(\mathrm{Frob}_v)) \in \mathfrak{m}$.

Conjecture A *Suppose that $\mathfrak{m}$ is non-Eisenstein and is associated to $\overline{\rho}$, and that Q is a set of primes v such that $N(v) \equiv 1 \mod p$, $\overline{\rho}$ is unramified at v, and $\overline{\rho}(\mathrm{Frob}_v)$ has distinct eigenvalues. Then there exists a continuous Galois representation $\rho = \rho_\mathfrak{m} : G_F \to \mathrm{GL}_2(\mathbf{T}_{Q,\mathfrak{m}})$ with the following properties:*

(1) *If $\lambda \notin S(\overline{\rho}) \cup Q \cup \{v|p\}$ is a prime of F, then ρ is unramified at λ, and the characteristic polynomial of $\rho(\mathrm{Frob}_\lambda)$ is*

$$X^2 - T_\lambda X + N_{F/\mathbf{Q}}(\lambda) \in \mathbf{T}_{Q,\mathfrak{m}}[X].$$

(2) *If $v \in S(\overline{\rho})$, then:*
   (a) *If $\overline{\rho}|D_v$ is of type* **S**, *then $\rho|I_v$ is unipotent.*
   (b) *If $\overline{\rho}|D_v$ is of type* **P**, *so that $\overline{\rho}|I_v \cong \psi \oplus \psi^{-1}$, then $\rho|I_v \cong \langle\psi\rangle \oplus \langle\psi\rangle^{-1}$.*

(3) *If $v \in Q$, the operators T_α for $\alpha \in F_v^\times \subset \mathbf{A}_F^{\infty,\times}$ are invertible. Let ϕ denote the character of $D_v = \mathrm{Gal}(\overline{F}_v/F_v)$ which, by class field theory, is associated to the resulting homomorphism:*

$$F_v^\times \to \mathbf{T}_{Q,\mathfrak{m}}^\times$$

given by sending x to T_x. By assumption, the image of $\phi \mod \mathfrak{m}$ is unramified, and so factors through $F_v^\times/\mathcal{O}_v^\times \simeq \mathbf{Z}$, and so $\phi(\mathrm{Frob}_v) \mod \mathfrak{m}$ is well defined; assume that $\phi(\mathrm{Frob}_v) \not\equiv \pm 1 \mod \mathfrak{m}$. Then $\rho|D_v \sim \phi\epsilon \oplus \phi^{-1}$.

(4) *If $v|p$, then $\overline{\rho}|D_v$ is finite flat, and if $\overline{\rho}|D_v$ is ordinary, then $\rho|D_v$ is ordinary.*

Some form of this conjecture has been suspected to be true at least as far back as the investigations of F. Grunewald in the early 70's (see [59,60]). Related conjectures about the existence of $\overline{\rho}_\mathfrak{m}$ were made for $\mathrm{GL}(n)/\mathbf{Q}$ by Ash [61], and for $\mathrm{GL}(2)/F$ by Figueiredo [62]. Say that a deformation of ρ_Q of $\overline{\rho}$ is minimal outside Q if it arises from a quotient of the ring R_Q of Sect. 5.1.

Lemma 5.6 *Assume Conjecture A. Assume that there exists a maximal ideal $\mathfrak{m}$ of $\mathbf{T}_Q$ associated to $\overline{\rho}$. Suppose that Q consists entirely of Taylor–Wiles primes. Then there exists a representation: $\rho_Q : G_F \to \mathrm{GL}_2(\mathbf{T}_{Q,\mathfrak{m}})$ whose traces generate $\mathbf{T}_{Q,\mathfrak{m}}$ and such that ρ_Q is a minimal deformation of $\overline{\rho}$ outside Q with cyclotomic determinant.*

Proof By Conjecture A, the representation $\rho_Q := \rho_\mathfrak{m}$ to $\mathbf{T}_{Q,\mathfrak{m}}$ is such a representation. Moreover, assumption 3 above guarantees (by Hensel's Lemma) that the T_α for $\alpha|Q$ lie in the $\mathcal{O}$-subalgebra generated by traces. □

5.3.1 Properties of homology groups

Let $\mathfrak{m}_\emptyset$ denote a non-Eisenstein maximal ideal of $\mathbf{T}_\emptyset$. We have natural homomorphisms

$$\mathbf{T}_Q^{\mathrm{an}} \to \mathbf{T}^{\mathrm{an}} = \mathbf{T}_\emptyset, \qquad \mathbf{T}_Q^{\mathrm{an}} \hookrightarrow \mathbf{T}_Q$$

induced by the map $H_1(Y_1(Q), \mathcal{O}) \to H_1(Y, \mathcal{O})$ and by the natural inclusion. (The surjectivity of this map is an immediate consequence of the interpretation of these groups in terms of group cohomology and the fact that the abelianization of $\mathrm{PSL}_2(\mathbf{F}_x)$ is trivial for $N(x) > 3$). The ideal $\mathfrak{m}_\emptyset$ of $\mathbf{T}_\emptyset$ pulls back to an ideal of $\mathbf{T}_Q^{\mathrm{an}}$ which we also denote by $\mathfrak{m}_\emptyset$ in a slight abuse of notation. The ideal $\mathfrak{m}_\emptyset$ may give rise to multiple maximal ideals $\mathfrak{m}$ of $\mathbf{T}_Q$.

Remark 5.7 If $x \notin Q \cup S(\overline{\rho}) \cup \{v|p\}$ is prime, then there is an operator $T_x \in \mathbf{T}_Q^{\mathrm{an}}$. If $x \in Q$, then we let U_x denote the operator $U_x := T_{\pi_x}$, where π_x, by abuse of notation, is the adele which is trivial away from x and the uniformizer π_x at x. However, this operator is only well defined up to a diamond operator $\langle\alpha\rangle$, where $\alpha \in \mathcal{O}_x^\times \subset \mathbf{A}_F^{\infty,\times}$. On the other hand, by Conjecture A, the image of U_x modulo $\mathfrak{m}$ is well defined, because the associated character ϕ is unramified.

If $x \notin S(\overline{\rho}) \cup \{v|p\}$ is a prime of F such that $N_{F/\mathbf{Q}}(x) \equiv 1 \mod p$ and $\overline{\rho}(\mathrm{Frob}_x)$ has distinct eigenvalues, then the representation $\overline{\rho}|G_x$ does not admit ramified semistable deformations. The following lemma is the homological manifestation of this fact.

Lemma 5.8 *Suppose that for each $x \in Q$ we have that $N_{F/\mathbf{Q}}(x) \equiv 1 \mod p$ and that the polynomial $X^2 - T_x X + N_{F/\mathbf{Q}}(x) \in \mathbf{T}_{\emptyset}[X]$ has distinct eigenvalues modulo $\mathfrak{m}_{\emptyset}$. Let $\mathfrak{m}$ denote the maximal ideal of $\mathbf{T}_Q$ containing $\mathfrak{m}_{\emptyset}$ and $U_x - \alpha_x$ for some choice of root α_x of $X^2 - T_x X + 1 \mod \mathfrak{m}$ for each $x \in Q$. Then there is an isomorphism of $\mathbf{T}^{\mathrm{an}}_{Q,\mathfrak{m}_{\emptyset}}$-modules*

$$H_1(Y_0(Q), \mathcal{O})_{\mathfrak{m}} \xrightarrow{\sim} H_1(Y, \mathcal{O})_{\mathfrak{m}_{\emptyset}}.$$

Proof Note that, by the universal coefficient theorem, we have $H^1(Y, K/\mathcal{O}) = H_1(Y, \mathcal{O})^{\vee}$ (and similarly for $Y_0(Q)$). We proceed as in the proof of Lemma 3.5 to deduce that there is an isomorphism

$$H^1(Y_0(x), K/\mathcal{O})_{\mathfrak{m}} = H^1(Y, K/\mathcal{O})_{\mathfrak{m}_{\emptyset}} \oplus V.$$

In light of the universal coefficient theorem, it suffices to show that $V = 0$. The remainder of the proof now proceeds as in Lemma 3.5. □

There is a natural covering map $Y_1(Q) \to Y_0(Q)$ with Galois group

$$\Delta_Q := \prod_{x \in Q} (\mathcal{O}_F/x)^{\times}.$$

If μ is a finitely generated $\mathcal{O}[\Delta_Q]$-module, it gives rise to a local system on $Y_0(Q)$. Let $\mathbf{T}^{\mathrm{univ}}$ be the polynomial algebra generated analogously to the one in Sect. 3.2.3 by Hecke endomorphisms T_{α} for all α which are trivial at primes in $Q \cup S(\overline{\rho}) \cup \{v|p\}$ and by U_x for $x \in Q$ (see Remark 5.7). We have an action of $\mathbf{T}^{\mathrm{univ}}$ on the homology groups $H_i(Y_0(Q), \mu)$ and the Borel–Moore homology groups $H_i^{BM}(Y_0(Q), \mu)$. The ideal $\mathfrak{m}_{\emptyset}$ gives rise to a maximal ideal $\mathfrak{m}$ of $\mathbf{T}^{\mathrm{univ}}$ after a choice of eigenvalue mod $\mathfrak{m}$ for U_x for all x dividing Q.

We let Δ denote a quotient of Δ_Q and $Y_{\Delta}(Q) \to Y_0(Q)$ the corresponding Galois cover. Further suppose that Δ is a p-power order quotient of Δ_Q. Then $\mathcal{O}[\Delta]$ is a local ring. Note that by Shapiro's Lemma there is an isomorphism $H_1(Y_0(Q), \mathcal{O}[\Delta]) \cong H_1(Y_{\Delta}(Q), \mathcal{O})$.

Lemma 5.9 *Let μ be a finitely generated $\mathcal{O}[\Delta]$-module. Then:*

(1) $H_i(Y_0(Q), \mu)_{\mathfrak{m}} = (0)$ *for* $i = 0, 3$.
(2) *If μ is p-torsion free, then $H_2(Y_0(Q), \mu)_{\mathfrak{m}}$ is p-torsion free.*

(3) *For all i, we have an isomorphism*

$$H_i(Y_0(Q),\mu)_{\mathfrak{m}} \xrightarrow{\sim} H_i^{BM}(Y_0(Q),\mu)_{\mathfrak{m}}.$$

Proof Consider part (1). By Nakayama's Lemma, we reduce to the case when $\mu = k$. Yet $H_3(Y_0(Q), k) = 0$ and the action of Hecke operators on $H_0(Y_0(Q), k)$ (which preserve the connected components) is via the degree map, and this action is Eisenstein (in the sense that the only $\mathfrak{m}$ in the support of H_0 are Eisenstein). For part (2), since μ is $\mathcal{O}$-flat (by assumption), there is an exact sequence

$$0 \to \mu \to \mu \to \mu/\varpi \to 0.$$

Taking cohomology, localizing at $\mathfrak{m}$, and using the vanishing of $H_3(Y_0(Q),\mu)_{\mathfrak{m}}$ from part (1), we deduce that $H^2(Y_0(Q),\mu)_{\mathfrak{m}}[\varpi] = 0$, hence the result. For part (3), there is a long exact sequence

$$\cdots \to H_i(\partial Y_0(Q),\mu) \to H_i(Y_0(Q),\mu) \to H_i^{BM}(Y_0(Q),\mu)$$
$$\to H_{i-1}(\partial Y_0(Q),\mu) \to \cdots$$

from which we observe that it suffices to show that $H_i(\partial Y_0(Q),\mu)_{\mathfrak{m}}$ vanishes for all i. (The action of Hecke operators on the boundary is the obvious one coming from topological considerations. For an explicit exposition of the relevant details, see p. 107 of [63]). By Nakayama's Lemma, we once more reduce to the case when $\mu = k$. The cusps are given by tori (specifically, elliptic curves with CM by some order in $\mathcal{O}_F$), and since the cohomology with constant coefficients of tori is torsion free, the case when $\mu = k$ reduces to the case when $\mu = \mathcal{O}$ and then $\mu = K$. We claim that the action of $\mathbf{T}^{\text{univ}}$ on the homology of the cusps in characteristic zero given by a sum of algebraic Grossencharacters for the field F; such a representation is Eisenstein by class field theory. This follows from the work of [64]; an explicit reference is Sect. 2.10 of [65]. □

Proposition 5.10 *The $\mathcal{O}[\Delta]$-module $H_1(Y_0(Q),\mathcal{O}[\Delta])_{\mathfrak{m}} \cong H_1(Y_\Delta(Q),\mathcal{O})_{\mathfrak{m}}$ is balanced (in the sense of Definition 2.2).*

Proof The argument is almost identical to the proof of Proposition 3.8. Let M denote the module $H_1(Y_0(Q),\mathcal{O}[\Delta])_{\mathfrak{m}}$ and $S = \mathcal{O}[\Delta]$. Consider the exact sequence of S-modules (with trivial Δ-action):

$$0 \to \mathcal{O} \xrightarrow{\varpi} \mathcal{O} \to k \to 0$$

where ϖ denotes a uniformizer in $\mathcal{O}$. Tensoring this exact sequence over S with M, we obtain an exact sequence:

$$0 \to \mathrm{Tor}_1^S(M, \mathcal{O})/\varpi \to \mathrm{Tor}_1^S(M, k) \to M_\Delta \to M_\Delta \to M \otimes_S k \to 0.$$

Let r denote the $\mathcal{O}$-rank of M_Δ. Then this exact sequence tells us that

$$d_S(M) = \dim_k M \otimes_S k - \dim_k \mathrm{Tor}_1^S(M, k) = r - \dim_k \mathrm{Tor}_1^S(M, \mathcal{O})/\varpi.$$

We have a Hochschild–Serre spectral sequence

$$H_i(\Delta, H_j(Y_0(Q), S)) = \mathrm{Tor}_i^S(H_j(Y_0(Q), S), \mathcal{O}) \implies H_{i+j}(Y_0(Q), \mathcal{O}).$$

We obtain an action of $\mathbf{T}^{\mathrm{univ}}$ on the spectral sequence by essentially the same argument as that of Proposition 3.8. Localizing at $\mathfrak{m}$, and using the fact that $H_i(Y_0(Q), S)_\mathfrak{m} = (0)$ for $i = 0, 3$ by Lemma 5.9 (1), we obtain an exact sequence

$$(H_2(Y_0(Q), S)_\mathfrak{m})_\Delta \to H_2(Y_0(Q), \mathcal{O})_\mathfrak{m} \to \mathrm{Tor}_1^S(M, \mathcal{O}) \to 0.$$

To show that $d_S(M) \geq 0$, we see that it suffices to show that $H_2(Y_0(Q), \mathcal{O})_\mathfrak{m}$ is free of rank r as an $\mathcal{O}$-module. By Lemma 5.9 (2), it then suffices to show that $\dim_K H_2(Y_0(Q), K)_\mathfrak{m} = r$. Inverting ϖ and applying Hochschild–Serre again, we obtain isomorphisms

$$(H_i(Y_0(Q), S \otimes_\mathcal{O} K)_\mathfrak{m})_\Delta \xrightarrow{\sim} H_i(Y_0(Q), K)_\mathfrak{m}$$

for $i = 1, 2$. It follows that $r = \dim_K H_1(Y_0(Q), K)_\mathfrak{m}$. By Poincaré duality, we have

$$\dim_K H_2(Y_0(Q), K)_\mathfrak{m} = \dim_K H_1^{BM}(Y_0(Q), K)_\mathfrak{m}.$$

(Because we are working with PGL, the dual maximal ideal $\mathfrak{m}^*$ is identified with $\mathfrak{m}$). Finally, by Lemma 5.9 (3), we have

$$\dim_K H_1(Y_0(Q), K)_\mathfrak{m} = \dim_K H_1^{BM}(Y_0(Q), K)_\mathfrak{m},$$

as required. □

5.4 Modularity lifting

We now associate to $\overline{\rho}$ the ideal $\mathfrak{m}_\emptyset$ of $\mathbf{T}_\emptyset$ which is generated by $(\varpi, T_\lambda - \mathrm{Trace}(\overline{\rho}(\mathrm{Frob}_\lambda)))$ where λ ranges over all primes $\lambda \notin S(\overline{\rho}) \cup \{v|p\}$ of F.

We make the hypothesis that $\mathfrak{m}_\emptyset$ is a *proper* ideal of $\mathbf{T}_\emptyset$. In other words, we are assuming that $\overline{\rho}$ is 'modular' of minimal level and trivial weight. Since $\mathbf{T}_\emptyset/\mathfrak{m}_\emptyset \hookrightarrow k$ it follows that $\mathfrak{m}_\emptyset$ is maximal. Since $\overline{\rho}$ is absolutely irreducible, it follows by Chebotarev density that $\mathfrak{m}_\emptyset$ is non-Eisenstein.

We now assume that Conjecture A holds for $\mathfrak{m}_\emptyset$. In other words, there is a continuous Galois representation

$$\rho_{\mathfrak{m}} : G_F \to \mathrm{GL}_2(\mathbf{T}_{\mathfrak{m}_\emptyset})$$

satisfying the properties of Conjecture A. The definition of $\mathfrak{m}_\emptyset$ and the Chebotarev density theorem imply that $\rho_{\mathfrak{m}_\emptyset} \mod \mathfrak{m}_\emptyset$ is isomorphic to $\overline{\rho}$. Properties (1)–(4) of Conjecture A then imply that $\rho_{\mathfrak{m}_\emptyset}$ gives rise to a homomorphism

$$\varphi : R^{\min} \to \mathbf{T}_{\mathfrak{m}_\emptyset}$$

such that the universal deformation pushes forward to $\rho_{\mathfrak{m}_\emptyset}$. The following is the main result of this section.

Theorem 5.11 *If we make the following assumptions:*

(1) *the ideal $\mathfrak{m}_\emptyset$ is a proper ideal of $\mathbf{T}_\emptyset$, and*
(2) *Conjecture A holds for all Q,*

then the map $\varphi : R^{\min} \to \mathbf{T}_{\mathfrak{m}_\emptyset}$ is an isomorphism and $\mathbf{T}_{\emptyset,\mathfrak{m}_\emptyset}$ acts freely on $H_1(Y, \mathcal{O})_{\mathfrak{m}_\emptyset}$.

Proof By property (1) of Conjecture A, the map $\varphi : R^{\min} \to \mathbf{T}_{\emptyset,\mathfrak{m}_\emptyset}$ is surjective. To prove the theorem, it therefore suffices to show that $H_1(Y, \mathcal{O})_{\mathfrak{m}_\emptyset}$ is free over $R^{\min}$ (where we view $H_1(Y, \mathcal{O})_{\mathfrak{m}_\emptyset}$ as an $R^{\min}$-module via φ). To show this, we will apply Proposition 2.3.

We set $R = R^{\min}$ and $H = H_1(Y, \mathcal{O})_{\mathfrak{m}_\emptyset}$ and we define

$$q := \dim_k H^1_\emptyset(G_F, \mathrm{ad}^0\overline{\rho}).$$

Note that $q \geq 1$ by Proposition 5.2. As in Proposition 2.3, we set $\Delta_\infty = \mathbf{Z}_p^q$ and let $\Delta_N = (\mathbf{Z}/p^N\mathbf{Z})^q$ for each integer $N \geq 1$. We also let R_∞ denote the power series ring $\mathcal{O}[x_1, \ldots, x_{q-1}]$. It remains to show that conditions 4 and 5 of Proposition 2.3 are satisfied. For this we will use the existence of Taylor–Wiles primes together with the results established in Sect. 5.2.

For each integer $N \geq 1$, fix a set of primes Q_N of F satisfying the properties of Proposition 5.2. We can and do fix a surjection $\widetilde{\phi}_N : R_\infty \twoheadrightarrow R_{Q_N}$ for each $N \geq 1$. We let ϕ_N denote the composition of $\widetilde{\phi}_N$ with the natural surjection $R_{Q_N} \twoheadrightarrow R^{\min}$. Let

$$\Delta_{Q_N} = \prod_{x \in Q_N} (\mathcal{O}_F/x)^\times$$

and choose a surjection $\Delta_{Q_N} \twoheadrightarrow \Delta_N$. Let $Y_{\Delta_N}(Q_N) \to Y_0(Q_N)$ denote the corresponding Galois cover. We set $H_N := H_1(Y_{\Delta_N}(Q_N), \mathcal{O})_{\mathfrak{m}}$ where $\mathfrak{m}$ is the ideal of $\mathbf{T}_{Q_N}$ which contains $\mathfrak{m}_\emptyset$ and $U_x - \alpha_x$ for each $x \in Q$, for some choice of α_x. Then H_N is naturally an $\mathcal{O}[\Delta_N] = S_N$-module. Applying Conjecture A to $\mathbf{T}_{Q_N,\mathfrak{m}}$, we deduce the existence of a surjective homomorphism $R_{Q_N} \twoheadrightarrow \mathbf{T}_{Q_N,\mathfrak{m}}$. Since $\mathbf{T}_{Q_N,\mathfrak{m}}$ acts on H_N, we get an induced action of R_∞ on H_N (via $\widetilde{\phi}_N$ and the map $R_{Q_N} \twoheadrightarrow \mathbf{T}_{Q_N,\mathfrak{m}}$). We can therefore view H_N as a module over $R_\infty \otimes_{\mathcal{O}} S_N$. To apply Proposition 2.3, it remains to check points (5a)–(5c). We check these conditions one by one:

(a) The image of S_N in $\mathrm{End}_{\mathcal{O}}(H_N)$ is contained in the image of R_∞ by Conjecture A, because it is given by the image of the diamond operators. The second part of condition (5a) follows from Conjecture A part (3) (exactly as in the proof of Theorem 3.26).

(b) We have a Hochschild–Serre spectral sequence

$$\mathrm{Tor}_i^{S_N}(H_j(Y_{\Delta_N}(Q_N), \mathcal{O})_{\mathfrak{m}}) \implies H_{i+j}(Y_0(Q_N), \mathcal{O})_{\mathfrak{m}}.$$

Applying part (1) of Lemma 5.9, we see that $(H_N)_{\Delta_N} \cong H_1(Y_0(Q_N), \mathcal{O})_{\mathfrak{m}}$. Then, by Lemma 5.8 we see that $(H_N)_{\Delta_N} \cong H_1(Y, \mathcal{O})_{\mathfrak{m}_\emptyset} = H$, as required.

(c) H_N is finite over $\mathcal{O}$ and hence over S_N. Proposition 5.10 implies that $d_{S_N}(H_N) \geq 0$.

We may therefore apply Proposition 2.3 to deduce that H is free over R and the theorem follows. □

If $H_1(Y, \mathcal{O})_{\mathfrak{m}_\emptyset} \otimes \mathbf{Q} \neq 0$, then we may deduce that the multiplicity μ for H as a $\mathbf{T}_{\emptyset,\mathfrak{m}_\emptyset}$-module is one by multiplicity one for $\mathrm{PGL}(2)/F$. The proof also exhibits $\mathbf{T}_{\emptyset,\mathfrak{m}_\emptyset}$ as a quotient of a power series ring in $q-1$ variables by q elements. In particular, if $\dim(\mathbf{T}_{\emptyset,\mathfrak{m}_\emptyset}) = 0$, then $\mathbf{T}_{\emptyset,\mathfrak{m}_\emptyset}$ is a complete intersection. From these remarks we see that Theorem 1.3 follows from Theorem 5.11.

5.5 The distinction between GL and PGL

The reader may wonder why, when considering Galois representations over imaginary quadratic fields, we consider the group $\mathbf{G} = \mathrm{PGL}$ rather than GL. When $F = \mathbf{Q}$ or an imaginary quadratic field, the associated locally symmetric spaces are very similar (the same up to components), and working with PGL has the disadvantage of forcing the determinant to be cyclotomic rather than cyclotomic up to finite twist. The main reason we use PGL is related to an issue which arises (and was pointed out to us by the referee) when the class number of F is divisible by p. Suppose that $\overline{\rho}$ is a modular representation of level one, and suppose that the minimal fixed determinant deformation ring is $\mathcal{O}$. Then, if the class group of $\mathcal{O}_F$ is $\mathbf{Z}/p\mathbf{Z}$, the minimal Hecke ring $\mathbf{T}_{\mathfrak{m}}$ is expected to be of the form $\mathcal{O}[\mathbf{Z}/p\mathbf{Z}]$ rather than $\mathcal{O}$, and the map $R^{\min} \rightarrow \mathbf{T}_{\mathfrak{m}}$ will not be surjective. The issue is that the Hecke algebra even at minimal level sees the twists of the corresponding automorphic form by characters of the class group. If these characters have p-power order, they contribute to the localization of $\mathbf{T}$ at any maximal ideal $\mathfrak{m}$. This is analogous to what might happen classically if one has a representation $\overline{\rho}$ over $\mathbf{Q}$ of tame level N; one has to be careful in choosing a minimal level, since the Hecke algebra of $X_1(N)$ will contain spurious twists if $N-1$ is divisible by p. The latter issue is easily resolved by a careful choice of level structure at N, namely, replacing $X_1(N)$ by $X_H(N)$ which is the quotient of $X_1(N)$ by the p-Sylow subgroup of the group $(\mathbf{Z}/N\mathbf{Z})^{\times}$ of diamond operators. However, it is not possible to avoid the *class group* in this way by choosing appropriate level structure, because the level structure only sees ramification. One fix is to work with PGL, but there is another fix for imaginary quadratic fields F which we sketch now. The natural approach is to replace the spaces Y, $Y_0(Q)$ and $Y_1(Q)$ by their quotients by the group $\mathrm{Cl}_p(\mathcal{O}_F) := \mathrm{Cl}(\mathcal{O}_F) \otimes \mathbf{Z}_p$. For example, the natural level structure at Y admits a ring of diamond operators which act via an extension of $\mathrm{Cl}(\mathcal{O}_F)$ by a group of order prime to p, and hence there is a canonical splitting and thus a canonical quotient $Y/\mathrm{Cl}_p(\mathcal{O}_F)$ which gives the "correct" space. Note that, for p odd, the group $\mathrm{Cl}_p(\mathcal{O}_F)$ acts freely on the components, so this quotient is given explicitly by a subset of the connected components of Y. In the example above, the natural ring of Hecke operators $\mathbf{T}_{\mathfrak{m}}$ acting on Y (now generated by T_α such that the image of (α) in $\mathrm{Cl}(\mathcal{O}_F)$ has order prime to p) will be isomorphic to $\mathcal{O}$. This construction, however, is not as canonical as one would like. For example, the ring of diamond operators on $Y_1(Q)$ naturally acts through a group whose p-Sylow subgroup is $\mathrm{RCl}_p(Q) = \mathrm{RCl}(Q) \otimes \mathbf{Z}_p$, the ($p$-part of the) ray class group of conductor Q. This group surjects onto $\mathrm{Cl}_p(\mathcal{O}_F)$, but there is no natural section. It seems that the Taylor–Wiles method still applies as long as one restricts the set of Taylor–Wiles primes to $x \in Q$ such that the map

$$\mathrm{RCl}_p(Q) \rightarrow \mathrm{Cl}_p(\mathcal{O}_F)$$

splits. This imposes a further Chebotarev condition on the Taylor–Wiles primes $x \in Q$ which corresponds to x splitting completely in a metabelian extension of F. Explicitly, if $\mathfrak{a} \in \mathrm{Cl}_p(\mathcal{O}_F)$ has p-power order h, let $\mathfrak{a}^h = (\alpha)$. The necessary condition on x is that (assuming p is prime to the order of the unit group of F) that

$$\alpha^{\frac{N(x)-1}{p}} \equiv 1 \mod x,$$

or equivalently that x splits completely in $F(\alpha^{1/p}, \zeta_p)$. We ultimately decided, however, to impose the simplifying assumption that $\det(\rho)$ is cyclotomic, in part because the main example of interest concerns elliptic curves over F which naturally have cyclotomic determinant.

Remark 5.12 A different approach to modularity lifting for GL is to allow the determinant to vary, specifically, to fix the determinant only up to a character which is ramified only at the Taylor–Wiles primes in Q (and so not at $v|p$). This is possibly the most general way to proceed, although it requires working with $\ell_0 > 0$ even for $\mathrm{GL}(2)/F$ for totally real fields of degree $[F : \mathbf{Q}] > 1$. We give some indication of this method by considering the case of $\mathrm{GL}(1)$ in Sect. 8.2. The general case of $\mathrm{GL}(n)$ is then a fibre product of this argument with the fixed determinant arguments for $\mathrm{PGL}(n)$.

Remark 5.13 Our methods may easily be modified to prove an $R^{\min} = \mathbf{T}_{\mathfrak{m}}$ theorem for ordinary representations in weights other than weight zero (given the appropriate modification of Conjecture A). In weights which are not invariant under the Cartan involution (complex conjugation), one knows *a priori* for non-Eisenstein ideals $\mathfrak{m}$ that $\mathbf{T}_{\mathfrak{m}}$ is finite. Note that in this case it is sometimes possible to prove unconditionally that $R^{\min}[1/p] = 0$, see Theorem 1.4 of [66].

Remark 5.14 One technical tool that is conspicuously absent when $l_0 = 1$ is the technique of solvable base change. When proving modularity results for $\mathrm{GL}(2)$ over totally real fields, for example, one may pass to a finite solvable extension to avoid various technical issues, such as level lowering (see [67]). However, if F is an imaginary quadratic field, then every non-trivial extension H/F has at least two pairs of complex places, and the corresponding invariant $l_0 = \mathrm{rank}(G) - \mathrm{rank}(K)$ for $\mathrm{PGL}(2)/H$ is at least 2 (more precisely, it is equal to the number of complex places of H). This means that when $l_0 = 1$, our techniques are mostly confined to the approach used originally by Wiles, Taylor–Wiles, and Diamond [1,2,26].

Remark 5.15 Our techniques also apply to some other situations in which $l_0 = 1$ (the **Betti** case). One may, for example, consider 2-dimensional representations over a field F with one complex place. If $[F : \mathbf{Q}]$ is even, there exists an inner form for $\mathrm{GL}(2)/F$ which is compact at all real places of F, and the corresponding arithmetic quotient is a finite volume arithmetic hyperbolic manifold which is compact if $[F : \mathbf{Q}] > 2$. (If $[F : \mathbf{Q}]$ is odd, one would have to require that $\overline{\rho}$ be ramified with semi-stable reduction at at least one prime $\lambda \nmid p$). Nonetheless, we obtain minimal lifting theorems in these cases, modulo an analogue of Conjecture A. Similarly, our methods immediately produce minimal lifting theorems for $\mathrm{GL}(3)/\mathbf{Q}$, modulo an appropriate version of Conjecture A. Similarly, our methods should also apply to other situations in which π_∞ is a holomorphic limit of discrete series (the **Coherent** case). One case to consider would be odd ordinary irreducible Galois representations $\rho : G_F \to \mathrm{GL}_2(\overline{\mathbf{Q}}_p)$ of a totally real field F

which conjecturally arise from Hilbert modular forms exactly one of whose weights is one. Other examples of particular interest include the case in which $\rho : G_{\mathbf{Q}} \to \mathrm{GSp}_4(\mathbf{Q}_p)$ is the Galois representation associated to an abelian surface $A/\mathbf{Q}$, or $\rho : G_E \to \mathrm{GL}_3(\mathbf{Q}_p)$ is the Galois representation associated to a Picard Curve (see the appendix to [68]). We hope to return to these examples in future work.

Part 2. l_0 **arbitrary.**

In this second part of the paper, our main result is a conditional modularity lifting theorem for n-dimensional p-adic representations of the Galois group of an arbitrary number field. In this generality, we are forced to work in a situation where the automorphic forms in question occur in a range of cohomological degrees of arbitrary length l_0. We could have presented our arguments in both the coherent cohomology setting and the Betti cohomology setting, but for concreteness, we have decided to treat only the latter case in detail.

We now state our main (conditional) modularity lifting theorem; it will be used in Sect. 10 to prove Theorem 1.1. Let $\mathcal{O}$ denote the ring of integers in a finite extension of $\mathbf{Q}_p$, let ϖ be a uniformizer of $\mathcal{O}$, and let $\mathcal{O}/\varpi = k$ be the residue field. Recall that a representation $\mathrm{Gal}(\mathbf{C}/\mathbf{R}) \to \mathrm{GL}_n(\mathcal{O})$ is *odd* if the image of c has trace in $\{-1, 0, 1\}$, and let ϵ denote the cyclotomic character.

Theorem 5.16 *Assume Conjecture B. Let $F/\mathbf{Q}$ be an arbitrary number field, and n a positive integer. Let $p > n$ be unramified in F. Let*

$$r : G_F \to \mathrm{GL}_n(\mathcal{O})$$

be a continuous Galois representation unramified outside a finite set of primes. Denote the mod-ϖ reduction of r by $\overline{r} : G_F \to \mathrm{GL}_n(k)$. Suppose that

(1) *If $v|p$, the representation $r|D_v$ is crystalline.*
(2) *If $v|p$, then $\mathrm{gr}^i(r \otimes_{\mathbf{Z}_p} B_{\mathrm{DR}})^{D_v} = 0$ unless $i \in \{0, 1, \dots, n-1\}$, in which case it is free of rank 1 over $\mathcal{O} \otimes_{\mathbf{Z}_p} F_v$.*
(3) *The restriction of $\overline{r}$ to $F(\zeta_p)$ is absolutely irreducible, and the field $F(\mathrm{ad}^0(\overline{r}))$ does not contain $F(\zeta_p)$.*
(4) *In the terminology of [3], Definition 2.5.1, $\overline{r}$ is big.*
(5) *If $v|\infty$ is any real place of F, then $r|G_{F_v}$ is odd.*
(6) *If r is ramified at a prime x, then $r|I_x$ is unipotent. Moreover, if, furthermore, $\overline{r}$ is unramified at x, then $N(x) \equiv 1 \mod p$.*
(7) *The determinant of r is $\epsilon^{n(n-1)/2}$.*
(8) *Either:*
   (a) *There exists a cuspidal automorphic representation π_0 of $\mathrm{GL}_n(\mathbf{A}_F)$ such that: $\pi_{0,v}$ has trivial infinitesimal character for all $v|\infty$, good reduction at all $v|p$, and the p-adic Galois representation $r_p(\pi)$ both satisfies condition 6 and the identity $\overline{r}_p(\pi) = \overline{r}$.*

(b) *$\bar{r}$ is Serre modular of minimal level $N(\bar{r})$, and r is ramified only at primes which ramify in $\bar{r}$.*

Then r is modular, that is, there exists a regular algebraic cusp form π for $\mathrm{GL}_n(\mathbf{A}_F)$ *with trivial infinitesimal character such that $L(r,s) = L(\pi,s)$.*

This theorem will follow immediately from Theorem 9.19, proved below. As in Theorem 1.2, condition 8b is only a statement about the existence of a mod-p cohomology class of level $N(\bar{r})$, not the existence of a characteristic zero lift; this condition is the natural generalization of Serre's conjecture. On the other hand, the usual strategy for proving potential modularity usually proceeds by producing characteristic zero lifts which are not minimal, and thus condition 8a will be useful for applications. If conditions 1, 2, and 3 are satisfied, then conditions 5 and 6 are satisfied after a solvable extension which is unramified at p. Moreover, *if* $\bar{r}$ admits an automorphic lift with trivial infinitesimal character and good reduction at p, then condition 8a is also satisfied after a solvable extension which is unramified at p. Condition 8b, however, is not obviously preserved under cyclic base change.

Note that it will be obvious to the expert that our methods will allow for (conditional) generalizations of these theorems to other contexts (for example, varying the weight) but we have contented ourselves with the simplest possible statements necessary to deduce Theorem 1.1. We caution, however, that several techniques are *not* available in this case, in particular, the lifting techniques of Ramakrishna and Khare–Wintenberger require that $l_0 = 0$.

6 Some commutative algebra II

The general difficulty in proving that $R_\infty = \mathbf{T}_\infty$ is to show that there are *enough* modular Galois representations. If the cohomology we are interested in occurs in a range of degrees of length l_0, then we would like to show that in at least one of these degrees that the associated modules H_N (which are both Hecke modules and modules for the group rings $S_N := \mathcal{O}[(\mathbf{Z}/p^N\mathbf{Z})^q])$ compile, in a Taylor–Wiles patching process, to form a module of codimension l_0 over the completed group ring $S_\infty := \mathcal{O}[(\mathbf{Z}_p)^q]$. The problem then becomes to find a find a suitable notion of "codimension l_0" for modules over a local ring that

(1) is well behaved for non-reduced quotients of power series rings over $\mathcal{O}$ (like S_N),
(2) can be established for the spaces H_N in question,
(3) compiles well in a Taylor–Wiles system.

It turns out to be more effective to patch together a series of complexes D_N of length l_0 whose cohomology computes the cohomology of $\Gamma_1(Q_N)$

localized at $\mathfrak{m}$. The limit of these patched complexes will then turn out to be a length l_0 resolution of an associated patched module.

It will be useful to prove the following lemmas.

Lemma 6.1 *Let S be a Noetherian local ring. If N is an S-module with depth n, and $0 \neq M \subseteq N$, then $\dim(M) \geq n$.*

Proof Let $\mathfrak{p}$ be an associated prime of M (and hence of N). Then $\mathfrak{p}$ is the annihilator of some $0 \neq m \in M$, and it suffices to prove the result for M replaced by $mS \subset M$. On the other hand, for a Noetherian local ring, one has the inequality (see [69], Theorem 17.2)

$$n = \operatorname{depth}(N) \leq \min_{\operatorname{Ass}(N)} \dim S/\mathfrak{p} \leq \dim(M).$$

□

We deduce from this the following:

Lemma 6.2 *Let $l_0 \geq 0$ be an integer and let S be a Noetherian regular local ring of dimension $n \geq l_0$. Let P be a perfect complex of S-modules which is concentrated in degrees $0, \ldots, l_0$. Then $\operatorname{codim}(H^*(P)) \leq l_0$, and moreover, if equality occurs, then:*

(1) *P is a projective resolution of $H^{l_0}(P)$,*
(2) *$H^{l_0}(P)$ has depth $n - l_0$ and has projective dimension l_0.*

Proof Let $\delta^i : P^i \to P^{i+1}$ denote the differential and let $m \leq l_0$ denote the smallest integer such that $H^m(P) \neq 0$. Consider the complex:

$$P^0 \to P^1 \to \cdots \to P^m.$$

By assumption, this complex is exact until the final term, and thus it is a projective resolution of the S-module $K^m := P^m/\operatorname{Im}(\delta^{m-1})$. It follows that the projective dimension of K^m is $\leq m$. On the other hand, we see that

$$H^m(P) = \ker(\delta^m)/\operatorname{Im}(\delta^{m-1}) \subseteq K^m,$$

and thus

$$\operatorname{codim}(H^m(P)) = n - \dim(H^m(P)) \leq n - \operatorname{depth}(K^m) = \operatorname{proj.dim}(K^m) \leq m,$$

where the central inequality is Lemma 6.1, and the second equality is the Auslander–Buchsbaum formula.

Suppose that $\operatorname{codim}(H^*(P)) \geq l_0$. Then it follows from the argument above that the smallest m for which $H^m(P)$ is non-zero is $m = l_0$, that $\operatorname{codim}(H^{l_0}(P)) = l_0$, that P is a resolution of $H^{l_0}(P)$, and that $\operatorname{proj.dim}(H^{l_0}(P)) = l_0$, completing the argument. □

6.1 Patching

We establish in this section an abstract Taylor–Wiles style patching result which may be viewed as an analogue of Theorem 2.1 of [26] and Proposition 2.3, but also including refinements due to Kisin.

Theorem 6.3 *Let q and j be non-negative integers with $q + j \geq l_0$, and let $S_\infty = \mathcal{O}[(\mathbf{Z}_p)^q]$. For each integer $N \geq 0$, let $S_N := \mathcal{O}[\Delta_N]$ with $\Delta_N := (\mathbf{Z}/p^N\mathbf{Z})^q$. For each $M \geq N \geq 0$ and each ideal I of $\mathcal{O}$, we regard S_N/I (and in particular, $\mathcal{O}/I = S_0/I$) as a quotient of S_M via the quotient map $\Delta_M \twoheadrightarrow \Delta_N$ and reduction modulo I.*

(1) *Let R_∞ be an object of $\mathcal{C}_\mathcal{O}$ of Krull dimension $1 + j + q - l_0$.*
(2) *Let R be an object of $\mathcal{C}_\mathcal{O}$, and let H be an R-module.*
(3) *Let T be a complex of finite-dimensional k-vector spaces concentrated in degrees $0, \ldots, l_0$ together with a differential $d = 0$ and an isomorphism $H^{l_0}(T) \xrightarrow{\sim} H/\varpi$ of k-modules.*

Let $\mathcal{O}^\square = \mathcal{O}[z_1, \ldots, z_j]$ and for each $\mathcal{O}$-module or $\mathcal{O}$-algebra M, we let $M^\square := M \otimes_\mathcal{O} \mathcal{O}^\square$. For any $\mathcal{O}$-algebra A, we regard A as a quotient of $A^\square$ via the map sending each z_i to 0.

Suppose that, for each integer $N \geq 1$, D_N is a perfect complex of S_N/ϖ^N-modules with the following properties:

(a) *There is an isomorphism $D_N \otimes_{S_N} S_N/\mathfrak{m}_{S_N} \simeq T$.*
(b) *For each $M \geq N \geq 0$ with $M \geq 1$ and each $n \geq 1$, there is an action of R_∞ on the on the cohomology of the complex $D_M^\square \otimes_{S_M} S_N/\varpi^n$ that commutes with that of $S_M^\square$. If, in addition, $N \geq N' \geq 0$ and $n \geq n' \geq 1$, then the natural map $H^*(D_M^\square \otimes_{S_M} S_N/\varpi^n) \to H^*(D_M^\square \otimes_{S_M} S_{N'}/\varpi^{n'})$ is compatible with the R_∞-actions.*
(c) *For each $N \geq 1$, there is a surjective map $\phi_N : R_\infty \to R$, and for each $n \geq 1$ we are given an isomorphism*

$$H^{l_0}(D_N^\square \otimes_{S_N^\square} \mathcal{O}/\varpi^n) = H^{l_0}(D_N \otimes_{S_N} \mathcal{O}/\varpi^n) \simeq H/\varpi^n$$

of R_∞-modules where R_∞ acts on H/ϖ^n via ϕ_N. Moreover, these isomorphisms are compatible for fixed N and varying n.
(d) *For M, N and n as above, the image of $S_M^\square$ in $\mathrm{End}_\mathcal{O}(H^*(D_M^\square \otimes_{S_M} S_N/\varpi^n))$ is contained in the image of R_∞ and moreover, the image of the augmentation ideal of $S_N^\square$ (that is, the kernel of $S_N^\square \to \mathcal{O}$) is contained in the image of* $\ker(\phi_N)$.

Let $\mathfrak{a} \subset S_\infty^\square$ denote the kernel of the map $S_\infty^\square \to \mathcal{O}$ sending each element of $(\mathbf{Z}_p)^q$ to 1 and each z_i to 0. Then the following holds: there is a perfect complex P_∞ of finitely generated S_∞-modules concentrated in degrees $0, \ldots, l_0$ such that

(i) *The complex $P_\infty^\square$ is a projective resolution, of minimal length, of its top degree cohomology $H^{l_0}(P_\infty^\square)$.*
(ii) *There is an action of $R_\infty \widehat{\otimes}_{\mathcal{O}} S_\infty^\square$ on $H^{l_0}(P_\infty^\square)$ extending the action of $S_\infty^\square$ and such that $H^{l_0}(P_\infty^\square)$ is a finite R_∞-module.*
(iii) *The R_∞-depth of $H^{l_0}(P_\infty)$ is equal to $1 + j + q - l_0 (= \dim R_\infty)$.*
(iv) *There is a surjection $\phi_\infty : R_\infty \twoheadrightarrow R$ and an isomorphism $\psi_\infty : H^{l_0}(P_\infty^\square)/\mathfrak{a} \xrightarrow{\sim} H$ of R_∞-modules where R_∞ acts on H via ϕ_∞. Moreover, the image of $\mathfrak{a}$ in $\mathrm{End}(H^{l_0}(P_\infty^\square))$ is contained in that of $\ker(\phi_\infty)$.*

Proof For each $N \geq 1$, let $\mathfrak{a}_N$ denote the kernel of the natural surjection $S_\infty \twoheadrightarrow S_N$ and let $\mathfrak{b}_N$ denote the open ideal of $S_\infty^\square$ generated by ϖ^N, $\mathfrak{a}_N$ and $(z_1^N, \ldots, z_j^N)$. Choose a sequence of open ideals $(\mathfrak{d}_N)_{N \geq 1}$ of R such that

- $\mathfrak{d}_N \supset \mathfrak{d}_{N+1}$ for all $N \geq 1$;
- $\cap_{N \geq 1} \mathfrak{d}_N = (0)$;
- $\varpi^N R \subset \mathfrak{d}_N \subset \varpi^N R + \mathrm{Ann}_R(H)$ for all N.

(As in the proof of Theorem 2.3, one can take $\mathfrak{d}_N$ to be the ideal generated by ϖ^N and $\mathrm{Ann}_R(H)^N$).

Define a *patching datum of level N* to be a 3-tuple (ϕ, ψ, P) where

- $\phi : R_\infty \twoheadrightarrow R/\mathfrak{d}_N$ is a surjection in $\mathcal{C}_{\mathcal{O}}$;
- P is a perfect complex of $S_\infty/(\mathfrak{a}_N + \varpi^N)$-modules such that $P \otimes S_\infty/\mathfrak{m}_{S_\infty} \simeq T$;
- For each $N \geq N' \geq 0$, each $N \geq n' \geq 1$ and each ideal I of $\mathcal{O}^\square$ with $(z_1^N, \ldots, z_j^N) \subset I \subset (z_1, \ldots, z_j)$, the cohomology groups $H^i(P^\square \otimes_{S_\infty^\square} S_{N'}^\square/(I + \varpi^{n'}))$ carry an action of R_∞ that commutes with the action of $S_\infty^\square$ and these R_∞-actions are compatible for varying N', n' and I;
- $\psi : H^{l_0}(P^\square \otimes_{S_\infty^\square} S_\infty^\square/(\mathfrak{a} + \varpi^N)) \xrightarrow{\sim} H/\varpi^N H$ is an isomorphism of R_∞ modules (where R_∞ acts on $H/\varpi^N H$ via ϕ). (Note that ψ then gives rise to an isomorphism of R_∞-modules between $H^{l_0}(P^\square \otimes_{S_\infty^\square} S_\infty^\square/(\mathfrak{a} + \varpi^{n'}))$ and $H/\varpi^{n'} H$ for each $N \geq n' \geq 1$).

We say that two such 3-tuples (ϕ, ψ, P) and (ϕ', ψ', P') are isomorphic if $\phi = \phi'$ and there is an isomorphism of complexes $P \xrightarrow{\sim} P'$ of S_∞-modules inducing isomorphisms of $R_\infty \widehat{\otimes}_{\mathcal{O}} S_\infty$-modules on cohomology which are compatible with ψ and ψ' in degree l_0. We note that, up to isomorphism, there are finitely many patching data of level N. (This follows from the fact that R_∞ and S_∞ are topologically finitely generated, and that T is finite). If D is a patching datum of level N and $1 \leq N' \leq N$, then D gives rise to a patching datum of level N' in an obvious fashion. We denote this datum by D mod N'.

For each pair of integers (M, N) with $M \geq N \geq 1$, we define a patching datum $D_{M,N}$ of level N as follows: the statement of the proposition gives a homomorphism $\phi_M : R_\infty \twoheadrightarrow R$ and an S_M/ϖ^M-complex D_M. We take

- ϕ to be the composition $R_\infty \overset{\phi_M}{\twoheadrightarrow} R \twoheadrightarrow R/\mathfrak{d}_N$;
- P to be $D_M \otimes_{S_\infty} S_\infty/(\mathfrak{a}_N + \varpi^N) = D_M \otimes_{S_M} S_N/\varpi^N$;
- $\psi : H^{l^0}(P^\square \otimes_{S_\infty^\square} S_\infty^\square/(\mathfrak{a} + \varpi^N)) = H^{l_0}(D_M^\square \otimes_{S_M^\square} \mathcal{O}/\varpi^N) \overset{\sim}{\to} H/\varpi^N$ to be the given isomorphism.

To see that the third condition in the definition of a patching datum is satisfied, let I be an ideal of $\mathcal{O}^\square$ with $(z_1^N, \ldots, z_j^N) \subset I \subset (z_1, \ldots, z_j)$, and let $1 \leq n' \leq N$, $0 \leq N' \leq N$. Then we have

$$H^i(P^\square \otimes_{S_\infty^\square} S_{N'}^\square/(I + \varpi^{n'})) = H^i(D_M^\square \otimes_{S_M} S_{N'}/\varpi^{n'}) \otimes_{\mathcal{O}^\square} \mathcal{O}^\square/I,$$

and hence, by assumption (b), this space carries an R_∞-action that commutes with the $S_\infty^\square$-action and is compatible for varying I, N' and n'. Thus, $D_{M,N}$ is indeed a patching datum of level N.

Since there are finitely many patching data of each level $N \geq 1$, up to isomorphism, we can find a sequence of pairs $(M_i, N_i)_{i \geq 1}$ such that

- $M_i \geq N_i$, $M_{i+1} \geq M_i$, and $N_{i+1} \geq N_i$ for all i;
- $D_{M_{i+1},N_{i+1}}$ mod N_i is isomorphic to D_{M_i,N_i} for all $i \geq 1$.

For each $i \geq 1$, we write $D_{M_i,N_i} = (\phi_i, \psi_i, P_i)$ and we fix an isomorphism between $D_{M_{i+1},N_{i+1}}$ mod N_i and D_{M_i,N_i}. We define

- $\phi_\infty : R_\infty \twoheadrightarrow R$ to be the inverse limit of the ϕ_i;
- $P_\infty := \varprojlim_i P_i$ where each transition map is the composite of $P_{i+1} \twoheadrightarrow P_{i+1}/(\varpi^{N_i} + \mathfrak{a}_{N_i})$ with the isomorphism $P_{i+1}/(\varpi^{N_i} + \mathfrak{a}_{N_i}) \overset{\sim}{\to} P_i$ coming from the chosen isomorphism between $D_{M_{i+1},N_{i+1}}$ mod N_i and D_{M_i,N_i}.
- ψ_∞ to be the isomorphism of R_∞-modules $H^{l_0}(P_\infty^\square)/\mathfrak{a} = H^{l_0}(P_\infty^\square/\mathfrak{a}) \overset{\sim}{\to} H$ (where R_∞ acts on H via ϕ_∞) arising from the isomorphisms ψ_i.

Then P_∞ is a perfect complex of S_∞-modules concentrated in degrees $0, \ldots, l_0$ such that $H^*(P_\infty^\square)$ carries an action of $R_\infty \widehat{\otimes}_{\mathcal{O}} S_\infty^\square$ (extending the action of $S_\infty^\square$). The image of $S_\infty^\square$ in $\mathrm{End}_{\mathcal{O}}(H^*(P_\infty^\square))$ is contained in the image of R_∞. (Use assumption (d), and the fact that the image of R_∞ is closed in $\mathrm{End}_{\mathcal{O}}(H^*(P_\infty^\square))$ (with its profinite topology)). It follows that $H^i(P_\infty^\square)$ is a finite R_∞-module for each i. Moreover, since $S_\infty^\square$ is formally smooth over $\mathcal{O}$, we can and do choose a homomorphism $\imath : S_\infty^\square \to R_\infty$ in $\mathcal{C}_{\mathcal{O}}$, compatible with the actions of $S_\infty^\square$ and R_∞ on $H^*(P_\infty^\square)$.

Since $\dim_{S_\infty^\square}(H^*(P_\infty^\square)) = \dim_{R_\infty}(H^*(P_\infty^\square))$ and $\dim R_\infty = \dim S_\infty^\square - l_0$, we deduce that $H^*(P_\infty^\square)$ has codimension at *least* l_0 as an $S_\infty^\square$-module. By Lemma 6.2 (with $S = S_\infty^\square$ and $P = P_\infty^\square$) we deduce that $P_\infty^\square$ is a resolution of minimal length of $H^{l_0}(P_\infty^\square)$ and that

$$\mathrm{depth}_{S_\infty^\square}(H^{l_0}(P_\infty^\square)) = \dim(S_\infty^\square) - l_0 = 1 + j + q - l_0.$$

Finally note that the image of of $\mathfrak{a}$ in $\mathrm{End}(H^{l_0}(P_\infty^\square))$ is contained in that of $\ker(\phi_\infty)$ by assumption (d). □

Theorem 6.4 *Keep the notation of the previous theorem and suppose in addition that R_∞ is p-torsion free.*

(1) *If R_∞ is formally smooth over $\mathcal{O}$, so $R \simeq \mathcal{O}[x_1, \ldots, x_{q+j-l_0}]$, then H is a free R-module.*
(2) *If $R_\infty[1/p]$ is irreducible, then H is a nearly faithful R-module (in the terminology of [11]).*
(3) *More generally, H is nearly faithful as an R-module providing that every irreducible component of* $\mathrm{Spec}(R_\infty[1/p])$ *is in the support of $H^{l_0}(P_\infty)[1/p]$.*

Proof Suppose first of all that $R_\infty \simeq \mathcal{O}[x_1, \ldots, x_{j+q-l_0}]$. Since $\mathrm{depth}_{R_\infty}(H^{l_0}(P_\infty^\square)) = \dim R_\infty$, applying the Auslander–Buchsbaum formula again, we deduce that $H^{l_0}(P_\infty^\square)$ is free over R_∞. Let $\iota : S_\infty^\square \to R_\infty$ be as in the proof of Theorem 6.3. Then, the existence of the isomorphism $\psi_\infty : H^{l_0}(P_\infty^\square)/\mathfrak{a}H^{l_0}(P_\infty^\square) \xrightarrow{\sim} H$ tells us that $R_\infty/\iota(\mathfrak{a})R_\infty$ acts freely on H and hence $\ker(\phi_\infty) \subset \iota(\mathfrak{a})R_\infty$. On the other hand, the freeness of $H^{l_0}(P_\infty^\square)$ over R_∞ and the fact that the image of $\mathfrak{a}$ in $\mathrm{End}(H^{l_0}(P_\infty^\square))$ is contained in that of $\ker(\phi_\infty)$ imply that $\iota(\mathfrak{a})R_\infty \subset \ker(\phi_\infty)$. We deduce that $\iota(\mathfrak{a})R_\infty = \ker(\phi_\infty)$ and that R acts freely on H, as required.

For the remaining cases, to show that H is nearly faithful as an R-module, it suffices to show that $H^{l_0}(P_\infty^\square)$ is nearly faithful as an R_∞-module. To see this, suppose $H^{l_0}(P_\infty^\square)$ is nearly faithful as an R_∞-module. Then $H^{l_0}(P_\infty^\square)/\mathfrak{a} \cong H$ is nearly faithful as an $R_\infty/\iota(\mathfrak{a})R_\infty$-module. The action of R_∞ on this module also factors through $R_\infty/\ker(\phi_\infty) = R$. Thus we see that $R^{\mathrm{red}} \twoheadrightarrow (R_\infty/\iota(\mathfrak{a})R_\infty)^{\mathrm{red}}$ and it suffices to show this map is an isomorphism. However, the fact that $H^{l_0}(P_\infty^\square)$ is nearly faithful as an R_∞-module together with the fact that the image of $\mathfrak{a}$ in $\mathrm{End}(H^{l_0}(P_\infty^\square))$ is contained in that of $\ker(\phi_\infty)$ imply that

$$\iota(\mathfrak{a}) + N \subset \ker(\phi_\infty) + N$$

where N is the ideal of nilpotent elements in R_∞. From this it follows immediately that $(R_\infty/\iota(\mathfrak{a}))^{\mathrm{red}} \twoheadrightarrow R^{\mathrm{red}}$, as required.

Since R_∞ is p-torsion free, all its minimal primes have characteristic 0. Thus $H^{l_0}(P_\infty^\square)$ is nearly faithful as an R_∞-module if and only if each irreducible component of $\mathrm{Spec}(R_\infty[1/p])$ lies in the support of $H^{l_0}(P_\infty^\square)[1/p]$. Part (3) follows immediately. For part (2), note that since $\mathrm{depth}_{R_\infty}(H^{l_0}(P_\infty^\square)) = \dim R_\infty$, the support of $H^{l_0}(P_\infty^\square)$ is a union of irreducible components of $\mathrm{Spec}(R_\infty)$ of maximal dimension. Since $H^{l_0}(P_\infty^\square) \neq \{0\}$, the result follows. □

Remark 6.5 It follows from the proof of the previous theorem that for $H^{l_0}(P_\infty^\square)$ to be nearly faithful as an R_∞-module, it is necessary that R_∞ be equidimensional.

To implement the level-changing techniques of [11], we will need the following refinement of Theorem 6.3.

Proposition 6.6 *Let S_N and $\mathcal{O}^\square$ be as in Theorem 6.3. Suppose we are given two sets of data $(R_\infty^i, R^i, H^i, T^i, (D_N^i)_{N\geq 1}, (\phi_N^i)_{N\geq 1})_{i=1,2}$ satisfying assumptions (1)–(3) and (a)–(d) of Theorem 6.3, for $i = 1, 2$. Suppose also that we are given:*

- *isomorphisms of k-algebras*

$$R_\infty^1/\varpi \xrightarrow{\sim} R_\infty^2/\varpi$$
$$R^1/\varpi \xrightarrow{\sim} R^2/\varpi,$$

- *an isomorphism of $R^1/\varpi \xrightarrow{\sim} R^2/\varpi$-modules*

$$H^1/\varpi \xrightarrow{\sim} H^2/\varpi,$$

- *an isomorphism, for each $M \geq N \geq 0$, of S_N/ϖ-modules*

$$H^{l_0}(D_M^1 \otimes_{S_M} S_N/\varpi) \xrightarrow{\sim} H^{l_0}(D_M^2 \otimes_{S_M} S_N/\varpi)$$

which induces (after tensoring over $\mathcal{O}$ with $\mathcal{O}^\square$) an isomorphism of $R_\infty^1 \otimes_{\mathcal{O}} S_N^\square/\varpi \xrightarrow{\sim} R_\infty^2 \otimes_{\mathcal{O}} S_N^\square/\varpi$-modules

$$H^{l_0}((D_M^1)^\square \otimes_{S_M} S_N/\varpi) \xrightarrow{\sim} H^{l_0}((D_M^2)^\square \otimes_{S_M} S_N/\varpi)$$

such that for each $M \geq 1$ the square

$$\begin{array}{ccc} H^{l_0}(D_M^1 \otimes_{S_M} \mathcal{O}/\varpi) & \longrightarrow & H^{l_0}(D_M^2 \otimes_{S_M} \mathcal{O}/\varpi) \\ \downarrow & & \downarrow \\ H^1/\varpi & \longrightarrow & H^2/\varpi \end{array}$$

commutes. Then we can find complexes $P_\infty^{i,\square}$ *for* $i = 1, 2$ *satisfying conclusions (i)–(iv) of Theorem 6.3 as well as the following additional property:*

- *There is an isomorphism of* $R_\infty^1/\varpi \xrightarrow{\sim} R_\infty^2/\varpi$*-modules*

$$H^{l_0}(P_\infty^{1,\square})/\varpi \xrightarrow{\sim} H^{l_0}(P_\infty^{2,\square})/\varpi$$

such that the square

$$\begin{array}{ccc} H^{l_0}(P_\infty^{1,\square})/(\mathfrak{a}+\varpi) & \longrightarrow & H^{l_0}(P_\infty^{2,\square})/(\mathfrak{a}+\varpi) \\ \downarrow & & \downarrow \\ H^1/\varpi & \longrightarrow & H^2/\varpi \end{array}$$

commutes.

Proof This can be proved in much the same way as Theorem 6.3; we omit the details. □

In practice, we will apply Prop 6.6 in a situation where we are primarily interested in the collection of data indexed by $i = 1$. For the data indexed by $i = 2$, the ring $R_\infty^2[1/p]$ will be irreducible and hence H^2 will be a nearly faithful R^2-module by Theorem 6.4. Proposition 6.6 will then allow us to deduce that H^1 is a nearly faithful R^1-module, following the arguments of [11].

7 Existence of complexes

In this section, we prove the existence of the appropriate perfect complexes of length l_0 which are required for patching. In both cases—the Betti case or the coherent case—the setting is similar: we have a covering space $X_\Delta(Q) \to X_0(Q)$ of manifolds or an étale map $X_\Delta(Q) \to X_0(Q)$ of schemes over $\mathcal{O}$, each with covering group Δ, which is a finite abelian group of the form $(\mathbf{Z}/p^N\mathbf{Z})^q$. In both cases, the cohomology localized at a maximal ideal $\mathfrak{m}$ of the corresponding Hecke algebra $\mathbf{T}$ is assumed to vanish outside a range of length l_0. The key point is thus to construct complexes of the appropriate length whose size is bounded (in the sense of condition (a) of Theorem 6.3) independently of Q, so that one may apply our patching result.

7.1 The Betti case

We put ourselves in the following somewhat general situation. Let $X_0(Q)$ denote the locally symmetric space associated to a reductive group G over some number field F and a compact open subgroup $K_0(Q)$ of $G(\mathbf{A}_F^\infty)$. Similarly let

$X_\Delta(Q)$ be associated to a normal subgroup $K_\Delta(Q) \subset K_0(Q)$ with quotient Δ, a finite abelian group. In practice, Q will represent a set of Taylor–Wiles primes. See Sect. 9 for the specific compact open subgroups that we will choose when $G = \mathrm{PGL}(n)$. Let $R = \mathcal{O}[\Delta]$ and let $\mathfrak{a}$ denote the augmentation ideal of R. Recall that by a *perfect complex* of R-modules, we mean a bounded complex of finite free R-modules. If Y is a topological space, we let $C(Y)$ denote the complex of $\mathcal{O}$-valued singular chains on Y.

Lemma 7.1 *There exists a perfect complex of R-modules C together with a quasi-isomorphism*

$$C \to C(X_\Delta(Q))$$

of complexes of R-modules. In particular, we have isomorphisms:

$$H_*(C \otimes_{\mathcal{O}} \mathcal{O}/\varpi^N) = H_*(X_\Delta(Q), \mathcal{O}/\varpi^N)$$
$$H_*(C \otimes_R R/(\mathfrak{a} + \varpi^N)) = H_*(X_0(Q), \mathcal{O}/\varpi^N)$$

for all integers N.

Proof By [43, Sect. II.5 Lemma 1], there exists a perfect complex of R-modules C together with a quasi-isomorphism $C \to C(X_\Delta(Q))$ of complexes of R-modules. (Mumford only guarantees that the final term of the complex is flat, but since R is local, this final term is also free). Since C and $C(X_\Delta(Q))$ are bounded complexes of flat R-modules, we have

$$H_*(C \otimes_R A) \xrightarrow{\sim} H_*(C(X_\Delta(Q)) \otimes_R A)$$

for every R-algebra A by [43, Sect. II.5 Lemma 2]. Taking $A = R/\varpi^N$ gives the first isomorphism. For the second isomorphism, the fact that $X_\Delta(Q) \to X_0(Q)$ is a covering map with group Δ implies that

$$C(X_\Delta(Q)) \otimes_R R/\mathfrak{a} \xrightarrow{\sim} C(X_0(Q)).$$

Thus, taking $A = R/(\mathfrak{a} + \varpi^N)$ gives the second isomorphism. □

Let $\gamma \in G(\mathbf{A}^{\infty,p})$ with associated Hecke operator $T_\gamma : H_*(X_\Delta(Q), \mathcal{O}/\varpi^N) \to H_*(X_\Delta(Q), \mathcal{O}/\varpi^N)$ for $N \le \infty$ (where we define $\mathcal{O}/\varpi^\infty := \mathcal{O}$).

Lemma 7.2 *Let C be as in Lemma 7.1. Then the action of T_γ on homology may be lifted to a map*

$$T_\gamma : C \to C$$

of complexes of R-modules.

Proof Let $A = \mathcal{O}/\varpi^N$ for $N \le \infty$ and let $K \subset G(\mathbf{A}_F^\infty)$ be the compact open subgroup with $X_K = X_\Delta(Q)$ (this was called $K_\Delta(Q)$ above). Let $K' = \gamma K \gamma^{-1} \cap K$ and $K'' = K \cap \gamma^{-1} K \gamma$. Note that right multiplication by γ gives an isomorphism: $\gamma : X_{K'} \to X_{K''}$. The operator T_γ is equal, up to an invertible scalar which we may ignore, to the composition:

$$H_i(X_K, A) \stackrel{\mathrm{tr}}{\to} H_i(X_{K'}, A) \stackrel{\gamma_*}{\to} H_i i(X_{K''}, A) \to H_i(X_K, A),$$

where the first map is the transfer (or corestriction) map. Thus T_γ is induced by the corresponding composition of morphisms of complexes

$$C(X_K) \stackrel{\mathrm{tr}}{\to} C(X_{K''}) \stackrel{\gamma_*}{\to} C(X_{K'}) \to C(X_K),$$

(after tensoring over $\mathcal{O}$ with A). Denote this composition $\widetilde{T}_\gamma$. Restricting to C (by means of the quasi-isomorphism $C \to C(X_K)$ of Lemma 7.1), we obtain $\widetilde{T}_\gamma : C \to C(X_K)$ which also gives rise to T_γ on homology. We thus have a diagram

$$\begin{array}{ccc} C & \xrightarrow{\widetilde{T}_\gamma} & C(X_K) \\ & & \uparrow \\ & & C \end{array}$$

of complexes of R-modules with the vertical morphism being a quasi-isomorphism. Since C is perfect, the morphism $\widetilde{T}_\gamma$ can be lifted to a morphism $T_\gamma : C \to C$ making the diagram commute. (See [70, Tag 08FQ] for example). ☐

Let $\mathbf{T}$ denote a Hecke algebra generated over $\mathcal{O}$ by a collection of operators T_γ. Then, for any $T \in \mathbf{T}$, we we can express T as a polynomial in the operators T_γ and thus lift the action of T on homology to an endomorphism $T : C \to C$.

Lemma 7.3 *For $T \in \mathbf{T}$, let $C_T := \varinjlim T^n C$. Then C_T is a perfect complex of R-modules whose homology is*

$$\varinjlim T^n H_*(C).$$

Proof Since R is complete, the functor $M \to \varinjlim T^n M$ on finitely generated R-modules is exact and and $\varinjlim T^n M$ is in fact a direct summand of M (the other factor being the submodule of M on which T is topologically nilpotent). A direct summand of a projective module is projective, and the equality of homology follows from the exactness of the functor. ☐

We now assume that Δ is of p-power order. Thus $R = \mathcal{O}[\Delta]$ is local and we let $\mathfrak{m}_R$ be the maximal ideal of R.

Lemma 7.4 (Nakayama's Lemma for perfect complexes) *Let $T \in \mathbf{T}$ be as above. Then there exists a perfect complex of R-modules D which is quasi-isomorphic to C_T and such that*

$$\dim D_n/\mathfrak{m}_R = \dim H_n(C_T \otimes R/\mathfrak{m}_R) = \dim \varinjlim T^m H_n(X_0(Q), k)$$

for all n. Moreover, the length of D is at most l_0, where l_0 is the range of cohomology groups such that $\varinjlim T^n H^*(C)$ is non-zero.

Proof By Lemma 1 of Mumford ([43], Chap. II.5) again, one may find a perfect complex K quasi-isomorphic to C_T and such that K is bounded of length l_0. Assume that the differential d on D is non-zero modulo $\mathfrak{m}_R$ from degree $n+1$ to n. Then by Nakayama's Lemma, there exists a direct sum decomposition of perfect complexes of R-modules

$$K \simeq L \oplus J,$$

where J_i is zero for $i \neq n+1, n$ and $d : J_{n+1} \to J_n$ is an isomorphism of free rank 1 R-modules. Thus, L is also a perfect complex of R-modules which is quasi-isomorphic to K. Replacing K by L and using induction, we eventually arrive at a complex D so that d is zero modulo $\mathfrak{m}_R$, from which the equality of dimensions follows by Nakayama's Lemma. □

In practice, the Hecke algebra $\mathbf{T}$ will be of the form $\mathbf{T}^{\mathrm{an}}[U_x : x \in Q]$ where $\mathbf{T}^{\mathrm{an}}$ is the subalgebra generated by good Hecke operators away from Q and p. We say that two maximal ideals of $\mathbf{T}$ give rise to the same Galois representation if they contract to the same ideal of $\mathbf{T}^{\mathrm{an}}$. In practice, we will be interested in localizing the homology groups $H_*(X_\Delta(Q), \mathcal{O}/\varpi^N)$ at a particular maximal ideal $\mathfrak{m}$ of $\mathbf{T}$. The residue field of $\mathfrak{m}$ will be equal to k. In order to apply the above lemmas, we take the Hecke operator T to be

$$\prod_{x \in Q} P_x \circ \prod_{i \in \Omega} P_i,$$

where Ω, P_x and P_i are chosen as follows: for each of the finitely many maximal ideals $\mathfrak{n}$ of $\mathbf{T}$ which occurs in $H_*(X_\Delta(Q), k)$, choose an $\mathcal{O}$-algebra homomorphism $\phi_\mathfrak{n} : \mathbf{T} \to \overline{k}$ with kernel $\mathfrak{n}$. We let Ω index the collection of such maximal ideals $\mathfrak{n}$ of $\mathbf{T}$ which give rise to a Galois representation distinct from $\mathfrak{m}$. This is equivalent to $\phi_\mathfrak{n}$ and $\phi_\mathfrak{m}$ differing on $\mathbf{T}^{\mathrm{an}}$. Thus, for $i \in \Omega$ corresponding to a maximal ideal $\mathfrak{n}$, there exists a good Hecke operator T_i such that $\phi_{\mathfrak{n}}(T_i) \neq \phi_{\mathfrak{m}}(T_i)$. Let $F_i(T)$ denote the minimal polynomial over E of the Teichmuller lift of $\phi_{\mathfrak{n}}(T_i)$ to $\overline{E}$. By Hensel's Lemma, the element $\phi_{\mathfrak{m}}(T_i) \in k$ is not a root modulo ϖ of $F_i(T)$. Hence $P_i = F_i(T_i)$ is an element of $\mathfrak{n}$ but not of $\mathfrak{m}$. For the maximal ideals $\mathfrak{n}$ with the same Galois representation as $\mathfrak{m}$, for all $x|Q$ by construction there will be a projector P_x which commutes with the action of the diamond operators and cuts out the localization at $\mathfrak{m}$. (For example, if $G = \mathrm{GL}(2)$, then P_x can be taken to be $\lim(U_x - \beta_x)^{n!}$ for $x|Q$, where $U_x - \alpha_x \in \mathfrak{m}$ and α_x, β_x are arbitrary lifts of the (distinct) eigenvalues of $\overline{r}(\mathrm{Frob}_x)$ to $\mathcal{O}$). In particular, we have

$$\lim_{\rightarrow} T^n H_*(\Gamma_\Delta(Q_N), \mathcal{O}/\varpi^N) = H_*(\Gamma_\Delta(Q_N), \mathcal{O}/\varpi^N)_{\mathfrak{m}}.$$

Thus, letting D be as in Lemma 7.4 for this choice of T, we have that D is a perfect complex of R-modules such that:

- $D_n \neq 0$ if and only if $H_n(X_0(Q), k)_{\mathfrak{m}} \neq \{0\}$,
- $H_n(D \otimes_{\mathcal{O}} \mathcal{O}/\varpi^N) \cong H_n(X_\Delta(Q), \mathcal{O}/\varpi^N)_{\mathfrak{m}}$, for all n, N, and
- $H_n(D \otimes_R R/(\mathfrak{a} + \varpi^N)) \cong H_n(X_0(Q), \mathcal{O}/\varpi^N)_{\mathfrak{m}}$ for all n, N.

7.2 The Coherent case

We now explain how to prove the existence of appropriate complexes in the setting of coherent cohomology. The setting will be as follows. We will have an étale map $Y = X_\Delta(Q) \to X = X_0(Q)$ with Galois group Δ, a finite abelian p-group. The spaces X and Y will be proper and smooth over $\mathrm{Spec}(\mathcal{O})$; they will arise as integral models of the Shimura varieties associated to some reductive group G over a number field F and some compact open subgroups $K_\Delta(Q) \subset K_0(Q) \subset G(\mathbf{A}_F^\infty)$. In the case that these Shimura varieties are not compact, X and Y will be arithmetic toroidal compactifications, as constructed in [71]. We will be given an automorphic vector bundle $\mathscr{E}$ on Y (in the case of a toroidal compactification, this will either be a canonical or subcanonical extension) which pulls back to a bundle also denoted by $\mathscr{E}$ on X. We will be interested in producing a perfect complex of R/ϖ^n-modules computing

$$H_i(Y, \mathscr{E} \otimes \mathcal{O}/\varpi^n)_{\mathfrak{m}}$$

where $\mathfrak{m}$ is a maximal ideal of the Hecke algebra generated by 'good' Hecke operators at the unramified primes together with certain operators at the primes in the set of auxiliary Taylor–Wiles primes Q; here the homology group is defined as

$$H_i(Y, \mathscr{E} \otimes \mathcal{O}/\varpi^n) := H^i(Y, \mathscr{E}^* \otimes_{\mathcal{O}_Y} \omega_Y \otimes_{\mathcal{O}} \mathcal{O}/\varpi^n)^\vee$$

where ω_Y is the determinant of $\Omega^1_{Y/\mathcal{O}}$. The reader may wonder why we introduce here the non-standard concept of "coherent homology." The reason is a mixture of both the practical and the psychological. In both the **Betti** and **Coherent** case, the modules which patch are obtained by taking the Pontryagin duals of the non-zero cohomology group in *lowest* degree with coefficients in $E/\mathcal{O}$. In Betti cohomology, the groups $H^{q_0}(X, \mathscr{E}/\mathcal{O})^\vee_\mathfrak{m}$ may be identified with the non-zero homology group of highest degree with coefficients in $\mathcal{O}$, namely $H_{q_0+l_0}(X, \mathcal{O})_\mathfrak{m}$. This identification is essentially a consequence of Poincaré duality (the manifolds X have boundary, but the ideals $\mathfrak{m}$ are chosen specifically so that the cohomology of the boundary becomes trivial after localization at $\mathfrak{m}$). In the coherent case, our use of the terminology "homology" is thus to preserve the analogy, but also as a convenient shorthand for the necessary operation of taking coefficients of our sheaves in $\mathscr{E}/\mathcal{O}$ and then taking Pontryagin duals. One could also make these analogies more precise by comparing Poincaré duality to Serre and Verdier duality.

We begin with some commutative algebra.

Lemma 7.5 *Let P be an $\mathcal{O}$-module such that P is ϖ-torsion free. Then P/ϖ^n is free over $\mathcal{O}/\varpi^n$ for each n.*

Proof If $n = 1$, then P/ϖ is a module over a field $k = \mathcal{O}/\varpi$, and hence admits a basis $\{\overline{x}_\alpha\}$. Let $\{x_\alpha\}$ denote any lift of this generating set to P. Let $Q \subset P$ denote the $\mathcal{O}$-submodule generated by the x_α. We claim that Q surjects onto P/ϖ^n for all n, and moreover that P/ϖ^n is free on the images of the generators x_α of Q. We prove this by induction. It is true for $n = 1$ by construction. Suppose that $Q \to P/\varpi^n$ is surjective. Let $x \in P$, and consider the image of x in P/ϖ^{n+1}. After subtracting a suitable element of Q, we may assume that the image of x in P/ϖ^n is trivial. Hence we may write $x = \varpi^n y$ for some $y \in P$. The image of y in P/ϖ can be written as the image of an element of Q, and so $y = z + \varpi w$ for $z \in Q$ and $w \in P$. It follows that $x = \varpi^n z \mod \varpi^{n+1}$, and thus the image of x in P/ϖ^{n+1} is contained in the image of Q. It follows that $Q \to P/\varpi^{n+1}$ is surjective. Let us now show that that P/ϖ^{n+1} is free over $\mathcal{O}/\varpi^{n+1}$. Assume otherwise. Then there exists a relation of the form

$$\sum r_\alpha x_\alpha \equiv 0 \mod \varpi^{n+1} P.$$

Reducing this equation modulo ϖ, we deduce by construction that r_α is divisible by ϖ for all α. Yet, since P is ϖ-torsion free, any equality $\varpi x = \varpi y$ in P implies the equality $x = y$. Hence if we write $r_\alpha = \varpi s_\alpha$, we obtain a relation

$$\sum s_\alpha x_\alpha \equiv 0 \mod \varpi^n P.$$

By induction, we deduce that s_α is 0 in $\mathcal{O}/\varpi^n$, and hence r_α is trivial in $\mathcal{O}/\varpi^{n+1}$. In particular, P/ϖ^{n+1} is freely generated by the images of the x_α, completing the induction. □

(As pointed out by the referee, Lemma 7.5 is also an immediate consequence of the fact that P/ϖ^m is flat over $\mathcal{O}/\varpi^m$ and so automatically free because $\mathcal{O}/\varpi^m$ is an Artinian local ring).

In the following lemma, Δ may be any finite abelian group.

Lemma 7.6 *Let $R = \mathcal{O}[\Delta]$. If M is an R-module which is free over $\mathcal{O}/\varpi^n$ and M/ϖ is free over R/ϖ, then M is free over R/ϖ^n.*

Proof Let $\{\overline{y}_\alpha\}$ be an $R/\varpi = k[\Delta]$ basis for M/ϖ. Since R is a free $\mathcal{O}$-module, we may choose a finite basis $\{z_i\}$ for R over $\mathcal{O}$. Note that $\{\overline{z}_i\}$ is a basis for $k[\Delta]$ over k, and $\overline{z}_i \cdot \overline{y}_\alpha$ is a basis for M/ϖ over k. Lifting the elements $\overline{y}_\alpha$ to elements y_α of M, we see, as in the proof of the Lemma 7.5, that $z_i \cdot y_\alpha$ is a free basis for M as an $\mathcal{O}/\varpi^n$ module. We claim that y_α is a free basis for M as an R/ϖ^n-module. Assume otherwise, so that there is a relation

$$\sum r_\alpha y_\alpha = 0$$

with $r_\alpha \in R/\varpi^n$. We may uniquely write $r_\alpha = \sum s_{\alpha,i} z_i$ with $s_{\alpha,i}$ in $\mathcal{O}/\varpi^n$. We then deduce from the freeness of M over $\mathcal{O}/\varpi^n$ with $z_i y_\alpha$ as a basis that $s_{\alpha,i} = 0$ for all α and all i, and hence $r_\alpha = 0$. □

We now return to the situation described at the beginning of this section: $f : Y \to X$ is an étale map of smooth proper $\mathcal{O}$-schemes with Galois group Δ abelian of p-power order. Let $\mathscr{F}$ be a coherent sheaf of $\mathcal{O}_X$-modules which is ϖ-torsion free. Following Nakajima [72], we take an affine covering of X by affine schemes $\{U_\alpha\}$ (which are necessarily flat over $\mathrm{Spec}(\mathcal{O})$). We thus obtain a Čech complex D of $\mathcal{O}$-flat $\mathcal{O}[\Delta]$-modules computing $H^i(Y, f^*(\mathscr{F}))$. More precisely, the terms of D are direct sums of modules of the form $N \otimes_A B$, where $\mathrm{Spec}(A) \subset X$ is an intersection of U_α's with preimage $\mathrm{Spec}(B) \subset Y$ and $N = \Gamma(\mathrm{Spec}(A), \mathcal{F})$. Moreover, the complex $D \otimes \mathcal{O}/\varpi^n$ computes $H^i(Y, f^*(\mathscr{F}) \otimes \mathcal{O}/\varpi^n)$ for every n.

Lemma 7.7 *Let* $\mathrm{Spec}(B) \to \mathrm{Spec}(A)$ *be a finite Galois étale morphism of flat $\mathcal{O}$-algebras with Galois group Δ. Let R denote the local ring* $\mathcal{O}[\Delta]$. *Then, for any A-module N which is flat over $\mathcal{O}$, $(N \otimes_A B)/\varpi^n$ is a free R/ϖ^n-module for each n.*

Proof Let $M = N \otimes_A B$. Then

$$M/\varpi = (N \otimes_A B)/\varpi = N/\varpi \otimes_{A/\varpi} B/\varpi,$$

and thus M/ϖ is a projective $R/\varpi = k[\Delta]$-module by Lemma 1 of [72], and is thus free. (A theorem of Kaplansky implies that a projective module P over a local ring R is always free (see [73])). Note that M is flat over $\mathcal{O}$ since N and A are flat over $\mathcal{O}$ and B is flat over A. It follows from Lemma 7.5 that M/ϖ^n is free over $\mathcal{O}/\varpi^n$. By Lemma 7.6, we deduce that M/ϖ^n is free over R/ϖ^n. □

It follows that $D \otimes \mathcal{O}/\varpi^n$ is a bounded complex of projective R/ϖ^n-modules computing the cohomology groups $H^i(Y, f^*(\mathscr{F})/\varpi^n)$. We apply this as follows. Let $\mathscr{E}$ denote the automorphic vector bundle introduced above and take

$$\mathscr{F} = \mathscr{E}^* \otimes_{\mathcal{O}_X} \omega_X.$$

Applying Lemma 1 of [43], Ch.II.5 again, we may replace $D \otimes \mathcal{O}/\varpi^n$ by a quasi-isomorphic complex of R/ϖ^n-modules $C_n^\vee$ which is perfect. Then dualizing this latter complex, we obtain a (chain) complex C_n whose homology groups are

$$H_i(Y, \mathscr{E} \otimes \mathcal{O}/\varpi^n).$$

It remains to define an action of the Hecke algebra and cut out the localization at $\mathfrak{m}$.

The Hecke algebra will be generated by operators T_γ where $\gamma \in G(\mathbf{A}_F^{\infty,p})$. Let $L = K_\Delta(Q) \subset G(\mathbf{A}^\infty)$ be the compact open subgroup corresponding to $Y = X_\Delta(Q)$. Let $L^\gamma = \gamma L \gamma^{-1} \cap L$. Then by choosing suitable polyhedral cone decomposition data, there exists an arithmetic toroidal compactification Y^γ, proper and smooth over $\mathcal{O}$, of the Shimura variety of level L^γ together with maps $\pi_1, \pi_2 : Y^\gamma \to Y$ where π_1 is associated to the inclusion $L^\gamma \to L$ and π_2 is associated to right multiplication by γ on complex points. (See [71, Sect. 6.4.3]). By [74, Thm. 2.15(4)(c)] and the fact that all automorphic vector bundles are constructed from the Hodge bundle (see [18, Sect. 4.2]), there is an isomorphism

$$\phi : \pi_2^* \mathscr{E} \xrightarrow{\sim} \pi_1^* \mathscr{E}$$

of sheaves on Y^γ. To define the Hecke operator T_γ, we follow the approach of [75, p. 256] which avoids having to define the trace of π_1 on cohomology.

Let $A = \mathcal{O}/\varpi^n$ and let $\mathscr{F} = \mathscr{E}^* \otimes \omega_X$ be as above. If M is an $\mathcal{O}$-module or a sheaf of $\mathcal{O}$-modules on some space, we denote by M_A the tensor product $M \otimes_{\mathcal{O}} A$. By Verdier duality, the group $H^i(Y, (f^*\mathscr{F})_A)$ is Pontryagin dual to

$$H^{d-i}(Y, \mathcal{H}om_{\mathcal{O}_Y}(f^*\mathscr{F}, \omega_Y)_A) = H^{d-i}(Y, (f^*\mathscr{E})_A)$$

where d is the dimension of Y. Thus, to define an operator T_γ on $H^i(Y, (f^*\mathscr{F})_A)$, it suffices to define a pairing

$$H^i(Y, (f^*\mathscr{F})_A) \otimes_A H^{d-i}(Y, (f^*\mathscr{E})_A) \to A.$$

We define the pairing by sending $x \otimes y$ to

$$\mathrm{tr}(\pi_1^*(x) \cup \phi\pi_2^*(y))$$

where

- $\pi_1^*(x) \cup \phi\pi_2^*(y)$ denotes the image of the cup product of $\pi_1^*(x)$ and $\phi\pi_2^*(y)$ under the natural map

$$H^d(Y^\gamma, (\mathscr{E}^* \otimes \omega_{Y^\gamma} \otimes \mathscr{E})_A) \to H^d(Y^\gamma, (\omega_{Y^\gamma})_A),$$

  and
- tr denotes the trace isomorphism

$$H^d(Y^\gamma, (\omega_{Y^\gamma})_A) \to A.$$

This defines the action of T_γ on cohomology. We now want to lift the action of T_γ to an endomorphism the complex C_n introduced above.

Let $Y_A = Y \times_{\mathcal{O}} A$ and let $\pi : Y_A \to \mathrm{Spec} A$ be the structural morphism. Then, since $\mathscr{G} := (f^*\mathscr{F})_A$ is a Δ-equivariant sheaf on Y_A, we may regard $R\pi_*(\mathscr{G}) = R\mathrm{Hom}(\mathcal{O}_{Y_A}, \mathscr{G})$ as an object of the bounded derived category of $R = A[\Delta]$-modules $D^b(R)$. By Verdier duality, we have

$$R\mathrm{Hom}(R\mathrm{Hom}(\mathcal{O}_{Y_A}, \mathscr{G}), A[0]) = R\mathrm{Hom}(\mathscr{G}, \omega_{Y_A}[d])$$

Thus, we have an equality (of Hom's in the category $D^b(R)$):

$$\begin{aligned}&\mathrm{Hom}(R\mathrm{Hom}(\mathcal{O}_{Y_A}, \mathscr{G}), R\mathrm{Hom}(\mathcal{O}_{Y_A}, \mathscr{G}))\\ &\quad = \mathrm{Hom}(R\mathrm{Hom}(\mathcal{O}_{Y_A}, \mathscr{G}) \overset{L}{\otimes} R\mathrm{Hom}(\mathscr{G}, \omega_{Y_A}[d]), A[0]),\end{aligned}$$

and we define an element of $\widetilde{T}_\gamma$ of the right hand side by composing

- the pullback under $\pi_1^* \otimes (\phi \circ \pi_2^*)$,

$$\begin{aligned}&R\mathrm{Hom}(\mathcal{O}_{Y_A}, \mathscr{G}) \overset{L}{\otimes} R\mathrm{Hom}(\mathscr{G}, \omega_{Y_A}[d])\\ &\quad \longrightarrow R\mathrm{Hom}\left(\mathcal{O}_{Y_A^\gamma}, \pi_1^*\mathscr{G}\right) \overset{L}{\otimes} R\mathrm{Hom}\left(\pi_1^*\mathscr{G}, \omega_{Y_A^\gamma}[d]\right)\end{aligned}$$

- the composition morphism

$$R\mathrm{Hom}(\mathcal{O}_{Y_A^\gamma}, \pi_1^*\mathcal{G}) \overset{L}{\otimes} R\mathrm{Hom}(\pi_1^*\mathcal{G}, \omega_{Y_A^\gamma}[d]) \to R\mathrm{Hom}(\mathcal{O}_{Y_A^\gamma}, \omega_{Y_A^\gamma}[d]),$$

  and
- the duality isomorphism

$$R\mathrm{Hom}(\mathcal{O}_{Y_A^\gamma}, \omega_{Y_A^\gamma}[d]) \xrightarrow{\sim} A[0].$$

Then $\widetilde{T}_\gamma$ induces the Hecke operator

$$T_\gamma : H^i(Y_A, \mathcal{G}) \to H^i(Y_A, \mathcal{G})$$

defined above. If C_n is the complex introduced above, then it comes equipped with an isomorphism

$$C_n^\vee \xrightarrow{\sim} R\pi_*(\mathcal{G}) = R\mathrm{Hom}(\mathcal{O}_{Y_A}, \mathcal{G})$$

in $D^b(R)$ and we have a diagram:

$$\begin{array}{ccccc} C_n^\vee & \longrightarrow & R\mathrm{Hom}(\mathcal{O}_{Y_A}, \mathcal{G}) & \xrightarrow{\widetilde{T}_\gamma} & R\mathrm{Hom}(\mathcal{O}_{Y_A}, \mathcal{G}) \\ & & & & \uparrow \\ & & & & C_n^\vee \end{array}$$

Since $C_n^\vee$ is perfect, we can apply [70, Tag 08FQ] once again to lift $\widetilde{T}_\gamma$ to a morphism of complexes $T_\gamma : C_n^\vee \to C_n^\vee$ that induces the operator T_γ on cohomology.

Now that we can lift a given Hecke operator T to the complex $C_n^\vee$ (and hence to its dual C_n), we can show, exactly as in the previous section, that given a maximal ideal $\mathfrak{m}$ of the Hecke algebra, then for a judicious choice of Hecke operator T, the complex

$$\varinjlim_m T^m C_n$$

is a perfect complex of R-modules with homology equal to

$$H_i(Y, \mathcal{E}_A)_{\mathfrak{m}}.$$

8 Galois deformations

8.1 Outline of what remains to be done to prove Theorem 5.16

Given the patching result (Theorem 6.3) and the existence of complexes satisfying an appropriate boundedness condition (condition (a) of *ibid*)., to complete the argument consists of the following steps. First, consider the *minimal* case where there are no primes $x \nmid p$ such that $\rho|D_x$ is ramified when $\overline{r}|D_x$ is unramified. In this case, it suffices to construct a sequence Q_N of collections of q primes $N(x) \equiv 1 \mod p^N$ such that the corresponding Hecke rings $\mathbf{T}_{Q_N}$ are all quotients of a fixed patched global deformation ring R_∞ such that $R_\infty[1/p]$ is irreducible of the appropriate dimension. The usual Taylor–Wiles–Kisin method (with some modifications due to Thorne) exactly produces the desired sets Q_N, and the computation of R_∞ (which is naturally a power series over R_{loc}) is computed in the usual manner, except now its dimension is l_0 less than in the classical case. If one *assumes* vanishing of cohomology outside the expected range and also *assumes* that the Hecke action of cohomology in that range comes from Galois representations with the expected properties, then one may construct a series of complexes (as in the last section) which all have actions by R_∞, and then using Theorem 6.3 the desired conclusions follow as expected.

This leaves the case when there exist primes such that $\rho|I_x$ is unipotent, $N(x) \equiv 1 \mod p$, and yet $\overline{r}|D_x$ is unramified. Here one uses Taylor's trick [11] to avoid Ihara's lemma. The key calculation in Taylor's paper requires only that one has control over the depth of the patched module on which the ring R_∞ acts, as well as the structure mod ϖ of various local deformation rings. The required information concerning depth is exactly what one deduces in the proof of Theorem 6.3. The only difference in this setting is that the relevant dimension of R_∞ is l_0 less than the classical case, whereas the corresponding depth of $H^{l_0}(P_\infty)$ is also exactly l_0 less than the classical case—this means that the argument goes through as expected. We begin, however, by explaining our method in the case of one-dimensional representations, in order to demonstrate the method.

8.2 Modularity of one-dimensional representations

In this section, we apply our method to one-dimensional representations, that is, to the case when $G = \mathrm{GL}(1)/F$ for an arbitrary number field F. We will need to assume that F does not contain ζ_p. The arguments here are (ultimately) somewhat circular, but they exhibit all the various aspects of the general method. The invariant value of ℓ_0 for a field F of signature (r_1, r_2) will be $r_1 + r_2 - 1$.

Up to twist, there is only one residual Galois representation, namely, the trivial representation

$$\overline{r} : G_F \to \mathbf{F}_p^\times.$$

Minimal deformations of $\overline{r}$ consist of (everywhere) unramified representations. Note that $\mathrm{ad}(\overline{r}) = \mathbf{F}_p$ and $\mathrm{ad}(\overline{r})(1) = \mu_p$. Let us consider the dimensions of the associated Selmer group $H^1_L(F, \mathbf{F}_p)$ and the dual Selmer group $H^1_{L^*}(F, \mu_p)$. Recall from the Greenberg–Wiles formula that:

$$\frac{|H^1_L(F, \mathbf{F}_p)|}{|H^1_{L^*}(F, \mathbf{F}_p)|} = \frac{|H^0(F, \mathbf{F}_p)|}{|H^0(F, \mu_p)|} \prod_v \frac{|L_v|}{|H^0(F_v, \mathbf{F}_p)|}.$$

The possible local contributions come from $v|p$, $v|\infty$, and the H^0 term. We assume that $\zeta_p \notin F$.

(1) The contribution from $\dfrac{|H^0(F, \mathbf{F}_p)|}{|H^0(F, \mu_p)|}$ is p, because $\zeta_p \notin F$. Thus the contribution to $\dim H^1_L(F, \mathbf{F}_p) - \dim H^1_{L^*}(F, \mu_p)$ is 1.
(2) Let $v|\infty$, so $F_v = \mathbf{R}$ or $F_v = \mathbf{C}$. The groups $H^0(\mathbf{R}, \mathbf{F}_p)$ and $H^0(\mathbf{C}, \mathbf{F}_p)$ are both one-dimensional. Hence, the contribution to $\dim H^1_L(F, \mathbf{F}_p) - \dim H^1_{L^*}(F, \mu_p)$ at the infinite places is $-r_1 - r_2$.
(3) Let $v|p$. Let k_v be the residue field of F_v. The group $H^0(F_v, \mathbf{F}_p)$ is always 1-dimensional. The Selmer condition $L_v \subset H^1(F_v, \mathbf{F}_p)$ is defined to be the classes that are unramified when restricted to inertia. By inflation–restriction, there is a map:

$$0 \to H^1(\mathrm{Gal}(\overline{k_v}/k_v), \mathbf{F}_p) \to H^1(F_v, \mathbf{F}_p) \to H^1(I_v, \mathbf{F}_p).$$

Since $H^1(\mathrm{Gal}(\overline{k}_v/k_v), \mathbf{F}_p)$ is clearly one-dimensional, the contribution of these terms is $1 - 1 = 0$ for each $v|p$.

It follows that

$$\dim H^1_L(F, \mathbf{F}_p) - \dim H^1_{L^*}(F, \mu_p) = -(r_1 + r_2 - 1).$$

We can, in fact, deduce this equality directly by computing both terms via class field theory. The first group has dimension $\dim \mathrm{Cl}(F)/p$. For the second, recall that for $v|p$ the group L_v^* is one-dimensional and is dual to the unramified classes in $H^1(F_v, \mathbf{F}_p)$. This dual consists of classes which are finite flat. So we are interested in $H^1_{\mathrm{fppf}}(\mathcal{O}_F, \mu_p)$, which sits in the exact sequence:

$$0 \to \mathcal{O}_F^\times/\mathcal{O}_F^{\times p} \to H^1_{\mathrm{fppf}}(\mathcal{O}_F, \mu_p) \to \mathrm{Cl}(F)[p] \to 0$$

by Hilbert's theorem 90 (Proposition III 4.9 of [33]). Hence the difference in ranks is

$$\dim \mathcal{O}_F^{\times}/\mathcal{O}_F^{\times p} = r_1 + r_2 - 1,$$

as follows from Dirichlet's unit theorem and the assumption that $\zeta_p \notin F$.

Now let us consider the corresponding symmetric space. The natural space to consider is

$$Y := J_F/F^{\times} = F^{\times}\backslash \mathbf{A}_F^{\times}/UK_{\infty}$$

where U is the maximal compact subgroup of the finite adeles, and K_{∞} is the connected component of the identity of the maximal compact subgroup of $F^{\times} \otimes \mathbf{R}$. For a set Q of auxiliary primes, we would also like to consider the space

$$Y_Q = F^{\times}\backslash \mathbf{A}_F^{\times}/U_Q$$

where , for $v \in Q$ with $N(v) \equiv 1 \mod p^n$, one replaces $\mathcal{O}_v^{\times}$ by $\mathcal{O}_v^{\times p^n}$, the unique subgroup of index p^n. In the corresponding notation for $\mathrm{GL}(n)$, we have $Y = Y_0(Q)$ and $Y_Q = Y_1(Q)$. It is slightly more aesthetically pleasing to work with the compact part of this space:

$$X_Q := F^{\times}\backslash \mathbf{A}_F^{\times}/U_Q A_{\infty}^0,$$

where A_{∞}^0 is the identity component of the $\mathbf{R}$-points of the maximal $\mathbf{Q}$-split torus in the centre. Note that Y_Q is an $\mathbf{R}$-bundle over X_Q, and so, from the cohomological viewpoint, the extra factor of $\mathbf{R}$ does not change any of the cohomology groups. What, geometrically, is X_Q? The component group is the maximal quotient of the ray class group of conductor Q and exponent p^n. The fibres are coming from the infinite primes. The group K_{∞} is r_2 copies of S^1 coming from the complex places. The fibres of Y_Q are then exactly the connected components of the cokernel of the map

$$\mathcal{O}_F^{\times} \rightarrow \mathbf{R}^{\times r_1} \times \mathbf{C}^{\times r_2}/K_{\infty},$$

which is $(S^1)^{r_1+r_2-1} \times \mathbf{R}$. Passing to X_Q excises the factor of $\mathbf{R}$. Hence the components of X_Q consist simply of a product of circles. Let us examine the cohomology of X_Q. In degree zero, the cohomology is

$$\mathbf{Z}_p[\mathrm{RCl}(Q)/p^n],$$

by class field theory. Let $\pi_v \in \mathcal{O}_v$ be a uniformizer, and suppose that $(v, Q) = 1$. Then the Hecke operator T_π acts on the component groups via the image of $[\pi_v] \in \mathrm{RCl}(Q)$. For $v|Q$, we also have diamond operators for $\alpha \in \mathcal{O}_v^\times$. There is a corresponding Galois representation

$$\rho_Q : G_F \to \mathbf{T}_Q^\times,$$

which is exactly the Galois representation coming from class field theory; the diamond operators correspond, via the local Artin map, to the representation of the inertia subgroups at $v|Q$. If we localize at the maximal ideal $\mathfrak{m}$ of $\mathbf{T}_Q$ corresponding to $\overline{\rho}$, we obtain a deformation of $\overline{\rho}$ which is ramified only at Q. On the other hand, the action of $\mathbf{T}_Q$ on the higher cohomology groups simply propagates from H^0 using the Künneth formula, and so one obtains the same action on all cohomology groups. If R_Q denotes the deformation ring of $\overline{\rho}$ unramified outside Q, we obtain a surjection

$$R_Q \to \mathbf{T}_{Q,\mathfrak{m}},$$

moreover, both rings are naturally modules over the ring of diamond operators $\mathbf{Z}_p[\Delta_Q] = \mathbf{Z}_p[U/U_Q]$, acting on the left via Yoneda's lemma and local class field theory, and on the right via Hecke operators; and this action is the same.

In this setting, a Taylor–Wiles prime v is a prime $N(v) \equiv 1 \mod p^n$ which satisfies a Galois condition and an automorphic condition. The automorphic condition is that there is no extra cohomology when passing from X to $X_0(Q)$. Since $X = X_0(Q)$, this is tautologically true. The Galois condition is that we have to be able to choose $|Q| = \dim H^1_{L^*}(F, \mu_p) = \dim H^1_L(F, \mathbf{F}_p) + \ell_0$ primes which exactly annihilate the dual Selmer group, which is $H^1_{L^*}(F, \mu_p) = H^1_{\mathrm{fppf}}(F, \mu_p)$. This group sits inside $H^1(F, \mu_p) = F^\times/F^{\times p}$, by Hilbert's Theorem 90 in the classical version. By Kummer theory, the corresponding elements give rise to extensions $F(\alpha^{1/p}, \mu_p)$. What does it mean to annihilate this class by allowing ramification at a prime $N(v) \equiv 1 \mod p^n$? Allowing ramification at v in the Selmer group corresponds to assuming that the classes in the dual Selmer group split completely at v. That is, we want to choose primes v so that the class is non-trivial under the map

$$H^1(F, \mu_p) \to H^1(F_v, \mu_p) = H^1(F_v, \mathbf{F}_p).$$

Note that $\mu_p = \mathbf{F}_p$ as a $\mathrm{Gal}(\overline{F}_v/F_v)$-module if $N(v) \equiv 1 \mod p$. This amounts to asking that the element Frob_v in $\mathrm{Gal}(F(\alpha^{1/p}, \zeta_p)/F)$ is non-trivial, and that $N(v) \equiv 1 \mod p^n$. By the Chebotarev density theorem, this is possible unless $F(\alpha^{1/p}) \subset F(\zeta_{p^n})$. Note that $\alpha \in F$. This can happen if $\alpha = \zeta_p$. So we have to assume that $\zeta_p \notin F$, although we have already made this assumption previously. More generally, if $\alpha^{1/p} \in F(\zeta_{p^n})$, then, since $F(\zeta_{p^n})$ is Galois an abelian over F, it must be the case that $F(\alpha^{1/p})$ is also Galois an abelian over F. This implies that $F(\zeta_p) \subset F(\alpha^{1/p})$, and a consideration of degrees implies that $\zeta_p \in F$. Hence, if $\zeta_p \notin F$, we may choose suitable primes to annihilate the dual Selmer group.

The invariant $\ell_0 = r_1 + r_2 - 1$ exactly matches the dimensions of cohomology which are supported on $\mathfrak{m}$ and the number of elements in the dual Selmer group which have to be annihilated (the dimension of X_Q is equal to ℓ_0, so the cohomology vanishing results in degrees $> \ell_0$ are automatic). Our patching result then allows us to deduce that there is an isomorphism

$$R^{\mathrm{unr}} \simeq \mathbf{Z}_p[\mathrm{Cl}(\mathcal{O}_F) \otimes \mathbf{Z}_p],$$

and hence an isomorphism between the Galois group of the maximal unramified abelian p-power extension of F and the class group of F.

Remark 8.1 The circularity of this argument comes from the application of Greenberg–Wiles, which requires the full use of class field theory. Even though we only apply this theorem in the seemingly innocuous case of $M = \mathbf{F}_p$, in fact the general proof of the Euler characteristic formula reduces exactly to this case by inflation.

8.3 Higher dimensional Galois representations

In this section, we apply our methods to Galois representations of regular weight over number fields. When the relevant local deformation rings are all smooth, the argument is similar to the corresponding result for imaginary quadratic fields in Sect. 5 that corresponds to the case $n = 2$ and $l_0 = 1$. However, in order to prove non-minimal modularity theorems, it is necessary to consider non-smooth rings, following Kisin. The main reference for many of the local computations below is the paper [3].

8.4 The invariant l_0

Let F be a number field of signature (r_1, r_2). The invariants l_0 and q_0 may be defined explicitly by the following formula, where "rank" denotes rank over $\mathbf{R}$.

$$\begin{aligned} l_0 &:= r_1\,(\mathrm{rank}(\mathrm{SL}_n(\mathbf{R})) - \mathrm{rank}(\mathrm{SO}_n(\mathbf{R}))) + r_2\,(\mathrm{rank}(\mathrm{SL}_n(\mathbf{C})) - \mathrm{rank}(\mathrm{SU}_n(\mathbf{C}))) \\ &= \begin{cases} r_1\left(\dfrac{n-1}{2}\right) + r_2(n-1), & n \text{ odd}, \\ r_1\left(\dfrac{n-2}{2}\right) + r_2(n-1), & n \text{ even}. \end{cases} \end{aligned}$$

$$\begin{aligned}2q_0 + l_0 &:= r_1\left(\dim(\mathrm{SL}_n(\mathbf{R})) - \dim(\mathrm{SO}_n(\mathbf{R}))\right) + r_2\left(\dim(\mathrm{SL}_n(\mathbf{C})) - \dim(\mathrm{SU}_n(\mathbf{C}))\right)\\ &= r_1\left(n^2 - 1 - \frac{n(n-1)}{2}\right) + r_2\left(2(n^2-1) - (n^2-1)\right).\end{aligned}$$

The invariants l_0 and $2q_0 + l_0$ arise as follows: $2q_0 + l_0$ is the (real) dimension of the locally symmetric space associated to $\mathbf{G} := \mathrm{Res}_{F/\mathbf{Q}}(\mathrm{PGL}(n))$, and $[q_0, \ldots, q_0 + l_0]$ is the range such that cuspidal automorphic π for $\mathbf{G}$ which are tempered at ∞ contribute to cuspidal cohomology (see Theorem 6.7, VII, p. 226 of [9]). (In particular, q_0 is an integer).

Let V be a representation of G_F of dimension n over a field of characteristic different from 2, and assume that the action of G_{F_v} is *odd* for each $v|\infty$. Explicitly, this is the trivial condition for complex places, and for real places $v|\infty$ says that the action of complex conjugation $c_v \in G_{F_v}$ satisfies $\mathrm{Trace}(c_v) \in \{-1, 0, 1\}$. Then, via an elementary calculation, one has:

$$\sum_{v|\infty} \dim H^0(F_v, \mathrm{ad}^0(V)) = \begin{cases} r_1\left(\dfrac{n^2+1}{2} - 1\right) + r_2(n^2-1), & n \text{ odd},\\ r_1\left(\dfrac{n^2}{2} - 1\right) + r_2(n^2-1), & n \text{ even}.\end{cases}$$

Thus, in both cases we see that:

$$\sum_{v|\infty} \dim H^0(F_v, \mathrm{ad}^0(V)) = [F:\mathbf{Q}]\frac{n(n-1)}{2} + l_0. \tag{1}$$

8.5 Deformations of Galois representations

Let $p > n$ be a prime that is unramified in F and assume that $\mathrm{Frac}\, W(k)$ contains the image of every embedding $F \hookrightarrow \overline{\mathbf{Q}}_p$. Fix a continuous absolutely irreducible representation:

$$\overline{r} : G_F \to \mathrm{GL}_n(k).$$

We assume that:

- For each $v|p$, $\overline{r}|_{G_v}$ is Fontaine–Laffaille with weights $[0, 1, \ldots, n-1]$ for each $\tau : \mathcal{O}_F \to k$ factoring through $\mathcal{O}_{F_v}$.
- For each $v \nmid p$, $\overline{r}|_{G_v}$ has at worst unipotent ramification and $\mathbf{N}_{F/\mathbf{Q}}(v) \equiv 1 \mod p$ if $\overline{r}$ is ramified at v.
- The restriction $\overline{r}|_{G_v}$ is odd for each $v|\infty$.

We also fix a continuous character

$$\xi : G_F \to \mathcal{O}^\times$$

lifting $\det(\overline{r})$ and with

- $\xi|_{I_v} = \epsilon^{n(n-1)/2}$ for all $v|p$, and
- $\xi|_{I_v} = 1$ for all $v \nmid p$.

For example, we can simply take $\xi = \epsilon^{n(n-1)/2}$.

Let S_p denote the set of primes of F lying above p. Let R denote a finite set of primes of F disjoint from S_p that contains all primes at which $\overline{r}$ ramifies and is such that $\mathbf{N}_{F/\mathbf{Q}}(v) \equiv 1 \mod p$ for each $v \in R$. Let Q denote a finite set of primes of F disjoint from $R \cup S_p$. Finally, let $S = S_p \coprod R$ and $S_Q = S \coprod Q$. In what follows, R will consist of primes away from p where we allow ramification and Q will consist of Taylor–Wiles primes.

8.5.1 Local deformation rings

For $v|p$, let R_v denote the framed Fontaine–Laffaille $\mathcal{O}$-deformation ring with determinant $\xi|_{G_{F_v}}$ and τ-weights equal to $[0, 1, \ldots, n-1]$ for each $\tau : \mathcal{O}_F \hookrightarrow W(k)$. By [3] Proposition 2.4.3, R_v is formally smooth over $\mathcal{O}$ of relative dimension $n^2 - 1 + [F_v : \mathbf{Q}_p]n(n-1)/2$.

For each $v \in R$, choose a tuple $\chi_v = (\chi_{v,1}, \ldots, \chi_{v,n})$ of distinct characters

$$\chi_{v,i} : I_v \longrightarrow 1 + \mathfrak{m}_{\mathcal{O}} \subset \mathcal{O}^\times$$

such that $\prod_i \chi_{v,i}$ is trivial. We introduce the following framed deformation rings for each $v \in R$:

- Let R_v^1 denote the universal framed $\mathcal{O}$-deformation ring of $\overline{r}|_{G_v}$ corresponding to lifts of determinant ξ and with the property that each element $\sigma \in I_v$ has characteristic polynomial $(X-1)^n$.
- Let $R_v^{\chi_v}$ denote the universal framed $\mathcal{O}$-deformation ring of $\overline{r}|_{G_v}$ corresponding to lifts of determinant ξ and with the property that each element $\sigma \in I_v$ has characteristic polynomial $\prod_i (X - \chi_{v,i}(\sigma))$.

We let

$$R_{\mathrm{loc}}^1 := \left(\widehat{\bigotimes}_{v \in S_p} R_v\right) \widehat{\bigotimes} \left(\widehat{\bigotimes}_{v \in R} R_v^1\right)$$
$$R_{\mathrm{loc}}^\chi := \left(\widehat{\bigotimes}_{v \in S_p} R_v\right) \widehat{\bigotimes} \left(\widehat{\bigotimes}_{v \in R} R_v^{\chi_v}\right)$$

Lemma 8.2 *The rings R_{loc}^1 and R_{loc}^χ have the following properties:*

(1) *Each of R_{loc}^1 and R_{loc}^χ is p-torsion free and equidimensional of dimension*

$$1 + |S_p \cup R|(n^2-1) + [F : \mathbf{Q}]\frac{n(n-1)}{2}.$$

(2) *We have a natural isomorphism:*

$$R^1_{\mathrm{loc}}/\varpi \xrightarrow{\sim} R^\chi_{\mathrm{loc}}/\varpi.$$

(3) *The topological space* $\operatorname{Spec} R^\chi_{\mathrm{loc}}$ *is irreducible.*
(4) *Every irreducible component of* $\operatorname{Spec} R^1_{\mathrm{loc}}/\varpi$ *is contained in a unique irreducible component of* $\operatorname{Spec} R^1_{\mathrm{loc}}$.

Proof This follows from [76, Lemma 3.3] using [11, Proposition 3.1] and the properties of the Fontaine–Laffaille rings R_v recalled above. □

For each $v \in Q$, we assume that:

- $\overline{r}|_{G_v} \cong \overline{s}_v \oplus \overline{\psi}_v$ where $\overline{\psi}_v$ is a generalized eigenspace of Frobenius of dimension 1.

Moreover, we let $\mathcal{D}_v$ denote the deformation problem (in the sense of [3, Defn. 2.2.2]) consisting of lifts r of $\overline{r}|_{G_v}$ of determinant ξ and of the form $\rho \cong s_v \oplus \psi_v$ where s_v (resp. ψ_v) lifts $\overline{s}_v$ (resp. $\overline{\psi}_v$) and I_v acts via (possibly different) scalars on s_v and ψ_v. Let

$$L_v \subset H^1(G_v, \mathrm{ad}^0(\overline{r}))$$

denote the Selmer condition determined by all deformations of $\overline{r}|G_v$ to $k[\epsilon]/\epsilon^2$ of type $\mathcal{D}_v$. Then

$$\dim_k L_v - h^0(G_v, \mathrm{ad}^0(\overline{r})) = 1.$$

8.5.2 Global deformation rings

We now consider the following global deformation data:

$$\begin{aligned}\mathcal{S}_Q &= (\overline{r}, \mathcal{O}, S_Q, \xi, (\mathcal{D}_v)_{v \in S_p \cup Q}, (\mathcal{D}^1_v)_{v \in R})\\ \mathcal{S}^\chi_Q &= (\overline{r}, \mathcal{O}, S_Q, \xi, (\mathcal{D}_v)_{v \in S_p \cup Q}, (\mathcal{D}^\chi_v)_{v \in R}),\end{aligned}$$

where $\mathcal{D}_v$, $\mathcal{D}^1_v$ and $\mathcal{D}^\chi_v$ are the local deformation problems determined by the rings R_v, R^1_v and $R^{\chi_v}_v$ for $v \in S_p$ or $v \in R$. A deformation of $\overline{r}$ to an object of $\mathcal{C}_\mathcal{O}$ is said to be of type $\mathcal{S}_Q$ (resp. $\mathcal{S}^\chi_Q$) if:

(1) it is unramified outside S_Q;
(2) it is of determinant ξ;
(3) for each $v \in S_p \cup Q$, it restricts to a lifting of type $\mathcal{D}_v$ and for $v \in R$, to a lifting of type $\mathcal{D}^1_v$ (resp. $\mathcal{D}^\chi_v$).

If $Q = \emptyset$, we will denote $\mathcal{S}_Q$ and $\mathcal{S}_Q^\chi$ by $\mathcal{S}$ and $\mathcal{S}^\chi$. The functor from $\mathcal{C}_\mathcal{O}$ to Sets sending R to the set of deformations of type $\mathcal{S}_Q$ (resp. $\mathcal{S}_Q^\chi$) is represented by an object $R_{\mathcal{S}_Q}$ (resp. $R_{\mathcal{S}_Q^\chi}$).

We will also need to introduce framing. To this end, let

$$T = S_p \cup R.$$

Let $R_{\mathcal{S}_Q}^{\square_T}$ (resp. $R_{\mathcal{S}_Q^\chi}^{\square_T}$) denote the object of $\mathcal{C}_\mathcal{O}$ representing the functor sending R in $\mathcal{C}_\mathcal{O}$ to the set of deformations of $\overline{r}$ of type $\mathcal{S}_Q$ (resp. $\mathcal{S}_Q^\chi$) framed at each $v \in T$. (We refer to Definitions 2.2.1 and 2.2.7 of [3] for the notion of a framed deformation of a given type, replacing the group $\mathcal{G}_n$ of *op. cit.* with GL_n where appropriate).

We have natural maps

$$R_{\mathcal{S}_Q} \longrightarrow R_{\mathcal{S}_Q}^{\square_T}$$
$$R_{\mathrm{loc}}^1 \longrightarrow R_{\mathcal{S}_Q}^{\square_T}$$

coming from the obvious forgetful maps on deformation functors. Similar maps exist for the 'χ-versions' of these rings. The following lemma is immediate:

Lemma 8.3 *The map*

$$R_{\mathcal{S}_Q} \longrightarrow R_{\mathcal{S}_Q}^{\square_T}$$

is formally smooth of relative dimension $n^2|T| - 1$. The same statement holds for the corresponding rings of type $\mathcal{S}_Q^\chi$.

We now consider the map $R_{\mathrm{loc}}^1 \to R_{\mathcal{S}_Q}^{\square_T}$. For this, we will need to consider the following Selmer groups:

$$H^1_{\mathcal{L}(Q),T}(G_F, \mathrm{ad}^0\overline{r}) := \ker\left(H^1(G_{F,S_Q}, \mathrm{ad}^0\overline{r}) \to \bigoplus_{x\in T} H^1(G_x, \mathrm{ad}^0\overline{r}) \bigoplus \bigoplus_{x\in Q} H^1(G_x, \mathrm{ad}^0\overline{r})/L_x\right)$$

$$H^1_{\mathcal{L}(Q)^\perp,T}(G_F, \mathrm{ad}^0\overline{r}(1)) := \ker\left(H^1(G_{F,S_Q}, \mathrm{ad}^0\overline{r}(1)) \to \bigoplus_{x\in Q} H^1(G_x, \mathrm{ad}^0\overline{r}(1))/L_x^\perp\right).$$

Proposition 8.4 (1) *The ring* $R^{\square_T}_{\mathcal{S}_Q}$ *(resp.* $R^{\square_T}_{\mathcal{S}^\chi_Q}$*) is a quotient of a power series ring over* R^1_{loc} *(resp.* R^χ_{loc}*) in*

$$h^1_{\mathcal{L},T}(G_F, \mathrm{ad}^0\overline{r}) + \sum_{v \in T} h^0(G_v, \mathrm{ad}\overline{r}) - h^0(G_F, \mathrm{ad}\overline{r}) \text{ variables.}$$

(2) *We have*

$$\begin{aligned} h^1_{\mathcal{L},T}(G_F, \mathrm{ad}^0\overline{r}) = h^1_{\mathcal{L}^\perp,T}&(G_F, \mathrm{ad}^0\overline{r}(1)) + h^0(G_F, \mathrm{ad}^0\overline{r}) \\ &-h^0(G_F, \mathrm{ad}^0\overline{r}(1)) \\ &+\sum_{v \in Q}(\dim_k L_v - h^0(G_v, \mathrm{ad}^0\overline{r})) \\ &- \sum_{v \in T \cup \{v \mid \infty\}} h^0(G_v, \mathrm{ad}^0\overline{r}). \end{aligned}$$

Proof The first part follows from the argument of [52, Lemma 3.2.2] while the second follows from Poitou–Tate duality and the global Euler characteristic formula (c.f. the proof of [52, Proposition 3.2.5]). □

8.6 The numerical coincidence

By choosing a set of Taylor–Wiles primes Q to kill the dual Selmer group, one deduces the following.

Proposition 8.5 *Assume that* $\overline{r}(G_{F(\zeta_p)})$ *is big and let* $q \geq h^1_{\mathcal{L}^\perp,T}(G_F, \mathrm{ad}^0\overline{r}(1))$ *be an integer. Then for any* $N \geq 1$*, we can find a tuple* $\left(Q, (\overline{\psi}_v)_{v \in Q}\right)$ *where*

(1) *Q is a finite set of primes of F disjoint from S with $|Q| = q$.*
(2) *For each $v \in Q$, we have $\overline{r}|_{G_v} \cong \overline{s}_v \oplus \overline{\psi}_v$ where $\overline{\psi}_v$ is a generalized eigenspace of Frobenius of dimension 1.*
(3) *For each $v \in Q$, we have $\mathbf{N}_{F/\mathbf{Q}}(v) \equiv 1 \mod p^N$.*
(4) *The ring* $R^{\square_T}_{\mathcal{S}_Q}$ *(resp.* $R^{\square_T}_{\mathcal{S}^\chi_Q}$*) is a quotient of a power series ring over* R^1_{loc} *(resp.* R^χ_{loc}*) in*

$$q + |T| - 1 - [F : \mathbf{Q}]\frac{n(n-1)}{2} - l_0.$$

variables.

Proof Suppose given a tuple $(Q, (\overline{\psi}_v)_{v \in Q})$ satisfying the first three properties. Let $e_v \in \mathrm{ad}\overline{r}$ denote the G_v-equivariant projection onto $\overline{\psi}_v$. Then, as in [3, Proposition 2.5.9] (although we work here with a slightly different deformation problem at each $v \in Q$) we have

$$0 \longrightarrow H^1_{\mathcal{L}(Q)^\perp,T}(G_F, \mathrm{ad}^0\overline{r}(1)) \longrightarrow H^1_{\mathcal{L}^\perp,T}(G_F, \mathrm{ad}^0\overline{r}(1)) \longrightarrow \oplus_{v\in Q} k$$

where the last map is given by $[\phi] \mapsto (\mathrm{tr}(e_v\phi(\mathrm{Frob}_v)))_v$. The argument of [3, Proposition 2.5.9] can then be applied to deduce that one may choose a tuple $(Q, (\overline{\psi}_x)_{x\in Q})$ satisfying the first three required properties and such that

$$H^1_{\mathcal{L}(Q)^\perp,T}(G_F, \mathrm{ad}^0\overline{r}(1)) = \{0\}.$$

The last property then follows from Proposition 8.4, equation (1) and the fact that

$$\dim_k L_v - h^0(G_v, \mathrm{ad}^0\overline{r}) = 1 \;\text{ if } v \in Q.$$

□

9 Homology of arithmetic quotients

Let $\mathbf{A}$ denote the adeles of $\mathbf{Q}$, and $\mathbf{A}^\infty$ the finite adeles. Similarly, let $\mathbf{A}_F$ and $\mathbf{A}_F^\infty$ denote the adeles and finite adeles of F. Let $\mathbf{G} = \mathrm{Res}_{F/\mathbf{Q}}\mathrm{PGL}(n)$, and write $G_\infty = \mathbf{G}(\mathbf{R}) = \mathrm{PGL}_n(\mathbf{R})^{r_1} \times \mathrm{PGL}_n(\mathbf{C})^{r_2}$. Let K_∞ denote a maximal compact of G_∞ with connected component K_∞^0. For any compact open subgroup K of $\mathbf{G}(\mathbf{A}^\infty)$, we may define an arithmetic orbifold $Y(K)$ as follows:

$$Y(K) := \mathbf{G}(\mathbf{Q})\backslash\mathbf{G}(\mathbf{A})/K_\infty^0 K.$$

It has dimension $2q_0 + l_0$ in the notation above. We will specifically be interested in the following K. Let $S = S_p \cup R$ and $S_Q = S \cup Q$ be as in Sect. 8. We follow the convention that the cohomology of $Y(K)$ is the orbifold cohomology in the sense of Remark 5.3.

9.1 Arithmetic quotients

Let $K_Q = \prod_v K_{Q,v}$ and $L_Q = \prod_v L_{Q,v}$ denote the open compact subgroups of $\mathbf{G}(\mathbf{A})$ such that:

(1) If $v \in Q$, $K_{Q,v}$ is the image in $\mathrm{PGL}(\mathcal{O}_v)$ of $\{g \in \mathrm{GL}_n(\mathcal{O}_v) \mid g \text{ stabilizes } \ell \bmod \pi_v\}$ where ℓ is a fixed line in k_v^n.

(2) If $v \in Q$, $L_{Q,v} \subset K_{Q,v}$ is the normal subgroup with quotient group $k_v^\times$. Explicitly,

$$K_{Q,v} = \begin{pmatrix} 1 & * \\ 0 & \mathrm{GL}_{N-1}(\mathcal{O}_v) \end{pmatrix} \quad \mathrm{mod}\ \pi_v,$$
$$L_{Q,v} = \begin{pmatrix} 1 & * \\ 0 & \mathrm{SL}_{N-1}(\mathcal{O}_v) \end{pmatrix} \quad \mathrm{mod}\ \pi_v.$$

(3) If $v \notin S_Q K_{Q,v} = L_{Q,v} = \mathrm{PGL}_n(\mathcal{O}_v)$.
(4) If $v \in R$, $K_{Q,v} = L_{Q,v} =$ the Iwahori $\mathrm{Iw}(v)$ subgroup of$\mathrm{PGL}_n(\mathcal{O}_v)$ associated to the upper triangular unipotent subgroup.

When $Q = \emptyset$, we let $Y = Y(K_\emptyset)$. Otherwise, we define the arithmetic quotients $Y_0(Q)$ and $Y_1(Q)$ to be $Y(K_Q)$ and $Y(L_Q)$ respectively. They are the analogues of the modular curves corresponding to the congruence subgroups consisting of $\Gamma_0(Q)$ and $\Gamma_1(Q)$. (Rather, they are the appropriate analogues in the PGL-context).

For each $v \in R$, let $\mathrm{Iw}_1(v) \subset \mathrm{Iw}(v)$ denote the pro-v Iwahori. We fix a character

$$\psi_v = \psi_{v,1} \times \cdots \times \psi_{v,n} : \mathrm{Iw}(v)/\mathrm{Iw}_1(v) \cong (k_v^\times)^n/(k_v^\times) \longrightarrow 1 + \mathfrak{m}_{\mathcal{O}} \subset \mathcal{O}^\times.$$

The collection of characters $\psi = (\psi_v)_{v \in R}$ allows us to define a local system of free rank 1 $\mathcal{O}$-modules $\mathcal{O}(\psi)$ on $Y_i(Q)$ for $i = 1, 2$: let $\pi : \widetilde{Y}_i(Q) \to Y_i(Q)$ denote the arithmetic quotient obtained by replacing the subgroup $\mathrm{Iw}(v)$ with $\mathrm{Iw}_1(v)$ for each $v \in R$. Then, a section of $\mathcal{O}(\psi)$ over an open subset $U \subset Y_i(Q)$ is a locally constant function $f : \pi^{-1}(U) \to \mathcal{O}$ such that $f(\gamma u) = \psi(\gamma) f(u)$ for all $\gamma \in \mathrm{Iw}(v)/\mathrm{Iw}_1(v)$. We let $H^i_\psi(Y_i(Q), \mathcal{O})$ and $H_{i,\psi}(Y_i(Q), \mathcal{O})$ denote $H^i(Y_i(Q), \mathcal{O}(\psi))$ and $H_i(Y_i(Q), \mathcal{O}(\psi))$. Note that if $\psi = 1$ is the collection of trivial characters, then $\mathcal{O}(\psi) \cong \mathcal{O}$ and hence $H^i_\psi(Y_i(Q), \mathcal{O}) \cong H^i(Y_i(Q), \mathcal{O})$.

9.2 Hecke operators

We recall the construction of the Hecke operators. Let $g \in \mathbf{G}(\mathbf{A}^\infty)$ be an invertible matrix trivial at each place $v \in R$. For $K \subset \mathbf{G}(\mathbf{A}^\infty)$ a compact open subgroup of the form K_Q or L_Q, the Hecke operator $T(g)$ is defined on the homology modules $H_{\bullet,\psi}(Y(K), \mathcal{O})$ by considering the composition:

$$\begin{aligned} H_{\bullet,\psi}(Y(K), \mathcal{O}) &\to H_{\bullet,\psi}(Y(gKg^{-1} \cap K_Q), \mathcal{O}) \\ &\to H_{\bullet,\psi}(Y(K \cap g^{-1}Kg), \mathcal{O}) \to H_{\bullet,\psi}(Y(K), \mathcal{O}), \end{aligned}$$

the first map coming from the corestriction map, the second coming from the map $Y(gKg^{-1} \cap K, \mathcal{O}) \to Y(K \cap g^{-1}Kg, \mathcal{O})$ induced by right multiplication by g on $\mathbf{G}(\mathbf{A})$ and the third coming from the natural map on homology. The maps on cohomology $H^{\bullet}_{\psi}(Y(K), \mathcal{O})$ are defined similarly. (Since, conjecturally, the cohomology of the boundary will vanish after localizing at the relevant $\mathfrak{m}$, we may work either with cohomology or homology, by duality). The Hecke operators act on $H^{\bullet}_{\psi}(Y(K), \mathcal{O})$ but do not preserve the homology of the connected components. For $\alpha \in \mathbf{A}_F^{\infty,\times}$, we define the Hecke operator $T_{\alpha,k}$ by taking

$$g = \operatorname{diag}(\alpha, \alpha, \ldots, 1, \ldots, 1)$$

consisting of k copies of α and $n - k$ copies of 1. We now define the Hecke algebra.

Definition 9.1 Let $\mathbf{T}^{\mathrm{an}}_{Q,\psi}$ denote the subring of

$$\operatorname{End} \bigoplus_{k,n} H^k_{\psi}(Y_1(Q), \mathcal{O}/\varpi^n)$$

generated by Hecke endomorphisms $T_{\alpha,k}$ for all $k \le n$ and all α which are units at primes in S_Q. Let $\mathbf{T}_{Q,\psi}$ denote the $\mathcal{O}$-algebra generated by the same operators with T_{α} for α non-trivial at places in Q. If $Q = \emptyset$, we write $\mathbf{T}_{\psi}$ for $\mathbf{T}_{Q,\psi}$.

If $\mathfrak{a} \subseteq \mathcal{O}_F$ is an ideal prime to the level, we may define the Hecke operator $T_{\mathfrak{a},k}$ as $(1/\mathbf{N}_{F/\mathbf{Q}}(\mathfrak{a})^k)T_{\alpha,k}$ where $\alpha \in \mathbf{A}_F^{\times,\infty}$ is any element which represents the ideal $\mathfrak{a}$ and such that α is 1 for each component dividing the level. In particular, if $\mathfrak{a} = x$ is prime, then T_x is uniquely defined when x is prime to the level but not when x divides the level.

Remark 9.2 It would be more typical to define $\mathbf{T}_{Q,\psi}$ as the subring of endomorphisms of

$$\operatorname{End} \bigoplus_{k} H^k_{\psi}(Y_1(Q), \mathcal{O}),$$

except that it would not be obvious from this definition that $\mathbf{T}_{Q,\psi}$ acts on $H^k_{\psi}(Y_1(Q), \mathcal{O}/\varpi^n)$ for any n. It may well be true (for the $\mathfrak{m}$ we consider) that $\mathbf{T}_{Q,\psi}$ acts *faithfully* on the module $H^{q_0+l_0}_{\psi}(Y_1(Q), \mathcal{O})$—and indeed (at least for $Q = \emptyset$) this (conjecturally) follows when Theorem 5.16 applies and (in addition) R_{loc} is smooth. Whether one can prove this directly is an interesting question. (The claim is obvious when the cohomology occurs in a range of length $l_0 = 0$, and also follows in the case $l_0 = 1$ given known facts about the action of $\mathbf{T}_Q$ on cohomology with K coefficients).

9.3 Conjectures on existence of Galois representations

Let $\mathfrak{m}$ denote a maximal ideal of $\mathbf{T}^{\mathrm{an}}_{Q,\psi}$, and let $\mathbf{T}^{\mathrm{an}}_{Q,\psi,\mathfrak{m}}$ denote the completion. It is a local ring which is finite (but not necessarily flat) over $\mathcal{O}$.

Conjecture B *There exists a semisimple continuous Galois representation* $\overline{r}_{\mathfrak{m}} : G_F \to \mathrm{GL}_n(\mathbf{T}^{\mathrm{an}}_{Q,\psi}/\mathfrak{m})$ *with the following property: if* $\lambda \notin S_Q$ *is a prime of* F*, then* $\overline{r}_{\mathfrak{m}}$ *is unramified at* λ*, and the characteristic polynomial of* $\overline{r}_{\mathfrak{m}}(\mathrm{Frob}_\lambda)$ *is*

$$X^n - T_{\lambda,1}X^{n-1} + \cdots + (-1)^i \mathbf{N}_{F/\mathbf{Q}}(\lambda)^{i(i-1)/2} T_{\lambda,i} X^{n-i} + \cdots + (-1)^n \mathbf{N}_{F/\mathbf{Q}}(\lambda)^{n(n-1)/2} T_{\lambda,n},$$

in $\mathbf{T}^{\mathrm{an}}_{Q,\psi}/\mathfrak{m}[X]$*. Note that this property determines* $\overline{r}_{\mathfrak{m}}$ *uniquely by the Chebotarev density theorem. If* $\overline{r}_{\mathfrak{m}}$ *is absolutely irreducible, we say that* $\mathfrak{m}$ *is* non-Eisenstein*. In this case we further predict that there exists a deformation* $r_{\mathfrak{m}} : G_F \to \mathrm{GL}_n(\mathbf{T}^{\mathrm{an}}_{Q,\psi,\mathfrak{m}})$ *of* $\overline{r}_{\mathfrak{m}}$ *unramified outside* S_Q *and such that the characteristic polynomial of* $r_{\mathfrak{m}}(\mathrm{Frob}_\lambda)$ *is given by the same formula as above. In addition, suppose that* $r_{\mathfrak{m}} \cong \overline{r}$ *(where* $\overline{r}$ *is the representation introduced in Sect. 8.5). Suppose also that the set of primes* Q *consists of a set of Taylor–Wiles primes, that is, a set of primes as constructed in Proposition 8.5; this is an empty condition when* $Q = \emptyset$*. We conjecture that* $r_{\mathfrak{m}}$ *enjoys the following properties:*

(1) *If* $v|p$*, then* $r_{\mathfrak{m}}|_{G_v}$ *is Fontaine–Laffaille with all weights equal to* $[0, 1, \ldots, n-1]$.
(2) *If* $v \in Q$*, then* $r_{\mathfrak{m}}|_{G_v}$ *is a lifting of type* $\mathcal{D}_v$ *where* $\mathcal{D}_v$ *is the local deformation problem specified in Sect. 8.5.1*.
(3) *If* $v \in R$*, then the characteristic polynomial of* $r_{\mathfrak{m}}(\sigma)$ *for each* $\sigma \in I_v$ *is* $(X - \psi_{v,1}(\mathrm{Art}_v^{-1}(\sigma)))\ldots(X - \psi_{v,n}(\mathrm{Art}_v^{-1}(\sigma)))$.
(4) (a) *The localizations* $H^i_{\psi}(Y_1(Q), \mathcal{O}/\varpi^m)_{\mathfrak{m}}$ *vanish unless* $i \in [q_0, \ldots, q_0 + l_0]$.
   (b) *The localizations* $H^i_{\psi}(\partial Y_1(Q), \mathcal{O}/\varpi^m)_{\mathfrak{m}}$ *vanish for all* i*, where* $\partial Y_1(Q)$ *is the boundary of the Borel–Serre compactification of* $Y_1(Q)$.
(5) *For* $x \in Q$*, let* $P_x(X) = (X - \alpha_x)Q_x(X)$ *denote the characteristic polynomial of* $\overline{r}(\mathrm{Frob}_x)$ *where* $\alpha_x = \overline{\psi}_x(\mathrm{Frob}_x)$*. Let* $\mathfrak{m}_Q$ *denote the maximal ideal of* $\mathbf{T}_{Q,\psi}$ *containing* $\mathfrak{m}$ *and* $V_x - \alpha_x$ *for all* $x|Q$*, where* $V_x = T_{x,1}$,

which is well defined modulo $\mathfrak{m}$. *Then there is an isomorphism*

$$\lim_{k\to\infty}\prod_{x\in Q} Q_x(V_x)^{k!} : H^*_{\psi}(Y,\mathcal{O}/\varpi^m)_{\mathfrak{m}} \xrightarrow{\sim} H^*_{\psi}(Y_0(Q),\mathcal{O}/\varpi^m)_{\mathfrak{m}_Q},$$

It follows that $r_{\mathfrak{m}}$ *is a deformation of* $\overline{r}$ *of type* $\mathcal{S}_Q$ *(resp.* $\mathcal{S}^{\chi}_Q$*) if each* ψ_v *is the trivial character (resp.* $\psi_v = \chi_v$ *for each* $v \in R$*). In this case, we obtain a surjection* $R_{\mathcal{S}_Q} \twoheadrightarrow \mathbf{T}^{\mathrm{an}}_{Q,1,\mathfrak{m}}$ $\left(\mathit{resp.}\ R_{\mathcal{S}^{\chi}_Q} \twoheadrightarrow \mathbf{T}^{\mathrm{an}}_{Q,\chi,\mathfrak{m}}\right)$.

Some form of this conjecture has been suspected to be true at least as far back as the investigations of F. Grunewald in the early 70's (see [59,60]). Related conjectures about the existence of $\overline{r}_{\mathfrak{m}}$ were made for $\mathrm{GL}(n)/\mathbf{Q}$ by Ash [61], and for $\mathrm{GL}(2)/F$ by Figueiredo [62]. One aspect of this conjecture is that it implies that the local properties of the (possibly torsion) Galois representations are captured by the characteristic zero local deformation rings $R^{\square}_v$ for primes v. One might hope that such a conjecture is true in maximal generality, but we feel comfortable making the conjecture in this case because the relevant local deformation rings (including the Fontaine–Laffaille deformation rings) reflect an honest integral theory, which is not necessarily true of all the local deformation rings constructed by Kisin, (although the work of Snowden [41] gives hope that at least in the ordinary case that local deformation rings may capture all integral phenomena). By dévissage, conditions 4 and 5 are satisfied if and only if they are satisfied for $n = 1$, e.g., with coefficients in the residue field $k = \mathcal{O}/\varpi$.

Remark 9.3 Part (4)(a) of Conjecture B may be verified directly in a number of small rank cases, in particular for $\mathrm{GL}(2)/F$ when F is CM field of degree either 2 or 4, or for $\mathrm{GL}(3)/\mathbf{Q}$. In the latter two cases (where $(q_0, l_0) = (2, 2)$ and $(2, 1)$ respectively), the key point is that the lattices in question satisfy the congruence subgroup property [77], which yields vanishing for both $(H^1)_{\mathfrak{m}}$ and (by duality, considering both $\mathfrak{m}$ and $\mathfrak{m}^*$) $(H^{q_0-1}_c)_{\mathfrak{m}}$ where $\mathfrak{m}$ is a non-Eisenstein maximal ideal. On the other hand, the vanishing of $(H^{q_0-1}_c)_{\mathfrak{m}}$ also implies the vanishing of $(H^{q_0-1})_{\mathfrak{m}}$ after localization at $\mathfrak{m}$, since the cohomology of the boundary vanishes after localization at $\mathfrak{m}$. (In these cases, we are implicitly using the fact that we know enough about the boundary of the locally symmetric varieties in question to also resolve Part (4)(b) of Conjecture B).

Remark 9.4 In stating Conjecture B, we have assumed that Q is divisible only by Taylor–Wiles primes. To modify the conjecture appropriately for more general Q, one would have to modify condition 2 to allow for more general quotients of the appropriate local deformation ring (which would involve a mix of tamely ramified principal series and unipotent representations) and one would also drop condition 5.

Remark 9.5 Condition 4 of Conjecture B says that we could also have formulated our conjecture for compactly supported cohomology, or equivalently for homology. The complexes we eventually patch are computing

$$H^*(Y, \mathcal{O}/\varpi^n)^\vee_{\mathfrak{m}} = H^*_c(Y, \mathcal{O}/\varpi^n)^\vee_{\mathfrak{m}} = H_*(Y, \mathcal{O}/\varpi^n)_{\mathfrak{m}^*}$$

for the dual maximal ideal $\mathfrak{m}$, so it may have made more sense to work with homology. Indeed, in the homological formulation, we wouldn't need to assume anything about the vanishing of the homology of the boundary localized at $\mathfrak{m}$. However, for historical reasons, we continue to work in the present setting, with the understanding that the real difficulty in Conjecture B lies (after Scholze [8]) with proving the non-vanishing of (co-)homology groups in the required range after localization at $\mathfrak{m}$ and proving local-global compatibility, especially at $v|p$.

The reason for condition 5 of Conjecture B is that the arguments of Sect. 3 of [3] (in particular, Lemma 3.2.2 of *ibid*). often require that the $\mathrm{GL}_n(F_x)$-modules M in question are $\mathcal{O}$-flat. However, it may be possible to remove this condition, we hope to return to this point later (it is also true that slightly weaker hypotheses are sufficient for our arguments). On the other hand, we have the following:

Lemma 9.6 *If $n = 2$, then condition 5 of Conjecture* B *holds for all Taylor–Wiles primes.*

Proof By induction, it suffices to prove the result when $Q = \{x\}$ consists of a single such prime. For simplicity, we drop ψ from the notation. The two natural degeneracy maps induce maps:

$$\phi : H^*(Y, \mathcal{O}/\varpi^n)^2 \to H^*(Y_0(x), \mathcal{O}/\varpi^n),$$
$$\phi^\vee : H^*(Y_0(x), \mathcal{O}/\varpi^n) \to H^*(Y, \mathcal{O}/\varpi^n)^2$$

such that the composition $\phi^\vee \circ \phi$ is the matrix

$$\begin{pmatrix} N(x)+1 & T_x \\ T_x & N(x)+1 \end{pmatrix},$$

which has determinant $T_x^2 - (1 + N(x))^2$. If the eigenvalues of $\overline{\rho}(\mathrm{Frob}_x)$ are α_x and β_x, then $\alpha_x\beta_x \equiv N(x) \equiv 1 \mod p$. If x is a Taylor–Wiles prime, then by assumption, α_x is distinct from β_x, or equivalently, $\alpha_x \not\equiv \pm 1 \mod p$. It follows that $T_x^2 - (1 + N(x))^2 \notin \mathfrak{m}$, and hence $\phi^\vee \circ \phi$ is invertible after localizing at $\mathfrak{m}$. In particular, the maps ϕ and $\phi^\vee$ induce a splitting

$$H^*(Y_0(x), \mathcal{O}/\varpi^n)_{\mathfrak{m}} \simeq H^*(Y, \mathcal{O}/\varpi^n)^2_{\mathfrak{m}} \oplus W,$$
$$H^*(Y_0(x), \mathcal{O}/\varpi^n)_{\widetilde{\mathfrak{m}}} \simeq H^*(Y, \mathcal{O}/\varpi^n)_{\widetilde{\mathfrak{m}}} \oplus W_{\widetilde{\mathfrak{m}}}$$

for some $\mathbf{T}^{\mathrm{an}}_{Q,\mathfrak{m}}$-module $W \subset H^*(Y_0(x), \mathcal{O}/\varpi^n)$, and $\widetilde{\mathfrak{m}} = (\mathfrak{m}, U_x - \alpha_x)$. (Here, by abuse of notation, α_x denotes any lift of $\alpha_x \in \mathcal{O}/\varpi$ to $\mathcal{O}/\varpi^n$). It suffices to prove that W is trivial. One approach is to try to show that some $w \in W$ generates either the Steinberg representation Sp or Sp $\otimes\, \chi$ for the quadratic unramified character χ of $F_x^\times$, and then deduce that the action of U_x on w is via ± 1, contradicting the assumption on $\mathfrak{m}$. However, the map

$$H^*(Y_0(x), \mathcal{O}/\varpi^n) \to \left(\lim_{\to} H^*(Y(x^m), \mathcal{O}/\varpi^n)\right)^{U_0(x)}$$

is *a priori* neither surjective nor injective, which causes some complications with this approach. Hence we proceed somewhat differently (although see Remark 9.8 below). The following argument is implicit in the discussion of Ihara's Lemma in Chapter 4 of [6].

Recall that there exists a decomposition

$$Y \simeq \coprod \Gamma_i \backslash \mathcal{H},$$

where the Γ_i are congruence subgroups commensurable with $\mathrm{PGL}_2(\mathcal{O}_F)$ and $\mathcal{H}$ denotes the corresponding locally symmetric space. The finitely many connected components of Y will naturally be a torsor over a ray class group corresponding to the level of Y. Let Γ be one such subgroup. Denote by Γ^1 the intersection $\Gamma \cap \mathrm{PSL}_2(\mathcal{O}_F)$. By construction, Γ/Γ^1 is an elementary two group, which we denote by Φ. Recall that we are assuming that $k = \mathcal{O}/\varpi$ has odd characteristic p. Then, by Hochschild–Serre, there is an isomorphism

$$H^*(\Gamma, \mathcal{O}/\varpi^n) \simeq H^*(\Gamma^1, \mathcal{O}/\varpi^n)^{\Phi}.$$

By construction, for a Taylor–Wiles prime x, the level structure of Y at x is maximal. For convenience, let us also assume that x is trivial in the ray class group corresponding to the component group of Y (this is equivalent to imposing a further congruence condition on x, but is imposed only for notational convenience in the argument below). For such a prime x, we may form the amalgam

$$G := \Gamma^1 *_{\Gamma^1_0(x)} \Gamma^1$$

of Γ^1 with itself along the subgroup $\Gamma^1_0(x) := \Gamma_0(x) \cap \Gamma^1$. Then G will be a congruence subgroup of the S-arithmetic group $\mathrm{PSL}_2(\mathcal{O}_F[1/x])$ with the same level structure of Γ^1 at primes away from x. (Without the extra assumption on x, one would have to amalgamate *different* pairs of lattices Γ_i occurring in the decomposition of Y according to the action of the ray class group, cf. Sect. 4.1.4 of [6]. The argument would then proceed quite similarly, but it would require more notation) The long exact sequence of Lyndon for an amalgam (See [78], p. 169) gives rise to the following exact sequence:

$$\cdots \to H^{i-1}(G, \mathcal{O}/\varpi^n) \to H^i(\Gamma^1, \mathcal{O}/\varpi^n)^2 \to H^i(\Gamma^1_0(x), \mathcal{O}/\varpi^n) \to \cdots$$

We claim that this sequence is equivariant with respect to the Hecke operator U_{x^2}, which acts on $H^*(G, \mathcal{O}/\varpi^n)$ by 1. First, recall the definition of U_{x^2}. It is defined to be $1/N(x)^2$ times the operator induced by taking the double coset operator for $\Gamma^1_0(x)$ corresponding to the matrix:

$$\begin{pmatrix} \pi_x^2 & 0 \\ 0 & 1 \end{pmatrix}.$$

Since this operator only depends on the matrix up to scalar, one may equally take the matrix to be

$$g := \begin{pmatrix} \pi_x & 0 \\ 0 & \pi_x^{-1} \end{pmatrix}.$$

With this normalization, the matrix g lies in G. In particular, the corresponding map on $H^*(G, \mathcal{O}/\varpi^n)$ is given by multiplication by the degree of this operator on $\Gamma^1_0(x)$, which is $N(x)^2$, and thus (after normalizing) it follows that U_{x^2} acts by 1. It follows that if we localize the sequence at any ideal $\mathfrak{m}$ such that $U_{x^2} - 1$ is invertible, then there is an isomorphism

$$H^i(\Gamma^1, \mathcal{O}/\varpi^n)^2_{\mathfrak{m}} \simeq H^i(\Gamma^1_0(x), \mathcal{O}/\varpi^n)_{\mathfrak{m}}.$$

To recover the isomorphism for Y, it suffices to repeat this argument for each lattice Γ_i. On $H^*(Y_0(x), \mathcal{O}/\varpi^n)$, however, the operator U_x satisfies $U_x^2 = U_{x^2}$. In particular, since neither α_x nor β_x is equal to ± 1, we deduce (for the maximal ideal $\mathfrak{m}$ of interest) that there is an isomorphism

$$H^i(Y, \mathcal{O}/\varpi^n)^2_{\mathfrak{m}} \simeq H^i(Y_0(x), \mathcal{O}/\varpi^n)_{\mathfrak{m}}.$$

Taking $\widetilde{\mathfrak{m}} = (\mathfrak{m}, U_x - \alpha_x)$ and applying the projections $\lim_{n\to\infty} (U_x - \beta_x)^n$ and $\lim_{n\to\infty} (U_x - \alpha_x)^n$ gives the necessary isomorphism. □

Remark 9.7 As noted in [6], the group $\mathrm{PSL}_2(F_x)$ decomposes as an amalgam whereas $\mathrm{PGL}_2(F_x)$ does not—this is the reason for the reduction step to the PSL_2 case above. One *could* proceed above with PGL_2, but then the amalgams would more naturally be subgroups of $\mathrm{PGL}_2(\mathcal{O}_F[1/x])^{(\mathrm{ev})} \subset \mathrm{PGL}_2(\mathcal{O}_F[1/x])$ consisting of matrices whose determinant has even valuation at x (cf. Chapter 4 of [6]). In either case, one deduces as above that U_x^2 acts by $+1$ on $H^*(G, \mathcal{O}/\varpi^m)$.

Remark 9.8 The proof of this lemma is related to the proof of Lemmas 3.5 and 5.8. However, in those cases, it was only necessary to prove equality in the lowest degree, which is more elementary. Indeed, if q_0 denotes the lowest degree in which $H^{q_0}(Y, \mathcal{O}/\varpi)_{\mathfrak{m}}$ is nonzero, then, by Hochschild–Serre, the kernel of the map

$$H^{q_0}(X_0(x), \mathcal{O}/\varpi^n) \to H^{q_0}(X_1(x), \mathcal{O}/\varpi^n)$$

has a filtration by terms of the form $H^i(\Delta, H^j(X_1(x), \mathcal{O}/\varpi^n))$ for $j < q_0$. Since (by assumption) the coefficients of this expression are trivial after localization at $\mathfrak{m}$ for $j < q_0$, the kernel vanishes and the map above is injective. Hence the argument above using the representation Π generated by $w \in W$ applies in this case. Analysis of this spectral sequence suggests, however, that the map localized at $\mathfrak{m}$ will not (in general) be injective for $i > q_0$ when $l_0 > 0$.

9.4 An approach to Conjecture B part 5

In this section, we present an informal approach to proving part 5 of Conjecture B under a stronger assumption that $\overline{r}$ has enormous image, at least in the analogous case of $\mathrm{GL}(n)$ (from which it should be easy to deduce the corresponding claim for PGL, since manifolds for the former are circle bundles over manifolds for the latter, and so have highly related Hecke actions). Here, by enormous image, we require (in addition to bigness) the existence of suitable Taylor–Wiles primes x such that $\overline{r}(\mathrm{Frob}_x)$ has *distinct* eigenvalues. We thank David Helm for some helpful remarks concerning the deformation theory of unramified principal series.

9.4.1 Local preliminaries

Let $F/\mathbf{Q}$ be a number field, let x be a prime in F such that $N(x) \equiv 1 \mod p$. Let $G = \mathrm{GL}_n(F_x)$, and let $D \subset B \subset P \subset G$ denote the Borel subgroup B and a parabolic subgroup P with Levi factor $L := \mathrm{GL}_{n-1}(F_x) \times F_x^\times$, and $D \simeq (F_x^\times)^n$ the Levi of B. Let $G(\mathcal{O}_x) = \mathrm{GL}_n(\mathcal{O}_x)$, let $U(x) \subset G(\mathcal{O}_x)$ denote the full congruence subgroup of level x, and let $U_0(x) \subset G(\mathcal{O}_x)$ denote the largest subgroup containing $U(x)$ whose image in $\mathrm{GL}_n(\mathcal{O}_x/\varpi_x)$ stabilizes a line, chosen compatibly with respect to P. Suppose that $N(x) \equiv 1 \mod p$.

Let $\overline{\rho} : G_x \to \mathrm{GL}_n(k)$ be a continuous semi-simple representation. We say that an irreducible admissible mod-p representation π is *associated* to $\overline{\rho}$ if

$$\mathrm{rec}(\pi) = \mathrm{WD}(\overline{\rho})$$

under the semi-simple local Langlands correspondence of Vignéras [79]. The following is well known.

Lemma 9.9 *Let $\overline{\rho} : G_x \to \mathrm{GL}_n(k)$ be unramified with distinct eigenvalues. Then $\mathrm{rec}(\pi) = \overline{\rho}$ if and only if π is the irreducible unramified mod-p principal series:*

$$\pi = \text{n-ind}_B^G(\chi),$$

where $\chi : (F_x^\times)^n \to k^\times$ factors through $(F_x^\times/\mathcal{O}_x^\times)^n$ and sends each uniformizer to a distinct eigenvalue of $\overline{\rho}(\mathrm{Frob}_x)$.

In addition, we have the following:

Lemma 9.10 *Let π be the unramified principal series in Lemma 9.9, and let π' denote any irreducible admissible representation of G such that $\pi' \not\simeq \pi$. Then $\mathrm{Ext}^1(\pi, \pi') = \mathrm{Ext}^1(\pi', \pi) = 0$.*

Proof The supercuspidal support of π consists of the distinct characters χ_i. If either extension group is non-zero, then, by Theorem 3.2.13 of [80], it follows that π' has the same supercuspidal support as π. But this implies that π' is a quotient of π, and hence is isomorphic to π. □

Definition 9.11 Let $\mathscr{C}$ denote the category of locally admissible G-modules over $A := \mathcal{O}/\varpi^k$ such that every irreducible subquotient of $M \in \mathscr{C}$ is associated to π.

Under our assumptions on $\overline{\rho}$, we may give a quite precise description of the finite length elements $M \in \mathscr{C}$.

Lemma 9.12 *Suppose that $M \in \mathscr{C}$ has finite length as a G-module. Then there exists a finite length A-module M_B and a character*

$$\widetilde{\chi} : B \to \mathrm{Aut}(M_B)$$

whose irreducible constituents correspond to the character χ, and such that $M \simeq \text{n-ind}_B^G(M_B)$.

Proof The irreducible constituents of the parabolic restriction $\mathrm{res}_G^B(\pi)$ consist of the characters χ^w for w in the Weyl group W of G. By assumption, all the eigenvalues of $\overline{\rho}(\mathrm{Frob}_x)$ are distinct, and hence all the characters χ^w are distinct. In particular, $\mathrm{Ext}^i(\chi^v, \chi^w) = 0$ for all i if $v \neq w \in W$. It follows that $\mathrm{res}_G^B(M)$ admits a decomposition

$$\mathrm{res}_G^B(M) \simeq \bigoplus_W M_B^w,$$

where the irreducible constituents of M_B^w are χ^w for $w \in W$. Moreover, $\mathrm{length}_A(M_B^w) = \mathrm{length}_G(M)$ is finite for any $w \in W$. Let $M_B := M_B^{\mathrm{id}}$. There is a natural map

$$M \to \text{n-ind}_B^G \mathrm{res}_G^B(M) = \bigoplus_W \text{n-ind}_B^G M_B^w \to \text{n-ind}_B^G M_B.$$

Note that M and $\text{n-ind}_B^G M_B$ are elements of $\mathscr{C}$ of the same length, and all the irreducible constituents of $\text{n-ind}_B^G M_B^w$ for $w \neq \mathrm{id}$ are distinct from π. Thus, by comparing lengths, to prove that the composite of these maps is an isomorphism it suffices to prove that the first map is injective. If K denotes the kernel, then $\mathrm{res}_G^B(K) = 0$. Yet this contradicts the assumption that M (and hence K) lies in $\mathscr{C}$, since $\mathrm{res}_G^B(\pi) \neq 0$. □

Recall that $D \subset B$ denotes the Levi of B, which is $(F^\times)^n$. Since χ is trivial on $D(\mathcal{O}_x)$, Any finite deformation $\widetilde{\chi}$ of $D(F)$ which deforms χ has pro-p image after restriction to $D(\mathcal{O}_x)$, and thus factors through

$$\left(\varprojlim F^\times/F^{\times p^m}\right)^n \simeq \left(\mathbf{Z}_p \oplus \varprojlim k^\times/k^{\times p^m}\right).$$

The universal deformation of this group can be given quite explicitly:

Corollary 9.13 *Let $q = |k^\times \otimes \mathbf{Z}_p|$. Then* n-ind_B^G *induces an equivalence of categories between the category of direct limits of finite length modules over the ring R below and and $\mathscr{C}$:*

$$R := \bigotimes_{i=1}^{n} A[T]/(T^q - 1) \otimes_{\mathcal{O}} A[X].$$

Using this description of $\mathscr{C}$, we may prove the following:

Lemma 9.14 *The category $\mathscr{C}$ has enough injectives. The functor $M \to M^{U(x)}$ from $\mathscr{C}$ to $G(k) := G(\mathcal{O}_x)/U(x) \simeq \mathrm{GL}_n(k)$-modules takes injectives to acyclic modules.*

Proof One may explicitly observe that the appropriate category of R-modules has enough injectives. The composite functor from R-modules to $G(k)$-modules can be described explicitly as follows. Given a deformation $\widetilde{\chi}$, recall that $\widetilde{\chi}$ factors through $(k^\times \times \mathbf{Z})^n$. Hence the restriction $\widetilde{\chi}|_{D(k)}$ to $(k^\times)^n = D(k) \subset B(k)$ is well defined, and one has

$$\left(\text{n-ind}_B^G(\widetilde{\chi})\right)^{U(x)} \simeq \mathrm{Ind}_{B(k)}^{G(k)}\left(\widetilde{\chi}|_{D(k)}\right).$$

Since finitely generated injective $A[T]/(T^q - 1)$-modules are free, it follows that the image of an injective module has a filtration whose pieces are isomorphic to $\Phi := \mathrm{Ind}_{B(k)}^{G(k)}(\psi)$, where ψ is the universal deformation over k of $D(k)$ to $k[k^\times/k^{\times q}]$. Yet Φ is a direct summand of $\mathrm{Ind}_{B(k)}^{G(k)} k[D(k)]$ and thus of $k[G(k)]$; hence it is injective and acyclic. □

Remark 9.15 Since the group $U(x)$ is pro-p, the higher cohomology of $U(x)$ vanishes. Hence, for any subgroup $U(x) \subset \Gamma \subset G(\mathcal{O}_x)$, by Hochschild–Serre there are identifications

$$H^i(\Gamma, M) \simeq H^i(\Gamma/U(x), M^{U(x)}).$$

Any injective $G(k)$-module is also injective as a $\Gamma/U(x) \subset G(k)$-module. Hence, by Lemma 9.14, the derived functors of $M \mapsto M^\Gamma$ are well defined, and they coincide with $H^i(\Gamma, M)$.

Suppose that the roots of $P(T) = (T - \alpha)Q(T)$ are the Satake parameters of π. The Hecke algebra of $U_0(x)$ contains the operator $V = V_{\varpi_x}$ corresponding to the double coset of the diagonal matrix with $n - 1$ entry 1 and the final entry ϖ_x. The operator $P(V)$ is zero on $\pi^{U_0(x)}$ and thus acts nilpotently on $M^{U_0(x)}$. In particular, if

$$e_\alpha := \varinjlim Q(V)^{n!},$$

then e_α is a projection of $M^{U_0(x)}$ onto the *localization* of $M^{U_0(x)}$ at the ideal $(V - \alpha)$ for any lift of α to $\mathcal{O}$.

We shall now define two functors $\mathscr{F}$ and $\mathscr{G}$ on the category $\mathscr{C}$, defined on objects by

$$\mathscr{F}(M) := M^{G(\mathcal{O}_x)}, \qquad \mathscr{G}(M) := e_\alpha M^{U_0(x)}.$$

There is a natural transformation: $\iota : \mathscr{F} \to \mathscr{G}$ defined by the composition of the obvious inclusion $M^{G(\mathcal{O}_x)} \hookrightarrow M^{U_0(x)}$ with e_α. Note that $\mathscr{F}$ and $\mathscr{G}$ are left exact. Since $\mathscr{C}$ has enough injectives (Lemma 9.14), we have associated right derived functors $R^k\mathscr{F}$ and $R^k\mathscr{G}$ respectively. Hence, since e_α is exact, we may (see Remark 9.15) identify these right derived functors with the following cohomology groups:

$$R^k\mathscr{F}(M) \simeq H^k(G(\mathcal{O}_x), M), \qquad R^k\mathscr{G}(M) \simeq e_\alpha H^k(U_0(x), M).$$

Theorem 9.16 *The natural transformation $\iota : \mathscr{F} \to \mathscr{G}$ is an isomorphism of functors. In particular, there is an isomorphism*

$$\iota_* : H^k(G(\mathcal{O}_x), M) \to e_\alpha H^k(U_0(x), M).$$

Remark 9.17 One should compare Theorem 9.16 to Lemma 3.2.2 of [3], which implies that ι_* is an isomorphism for $k = 0$ and modules $M \in \mathscr{C}$ of the form $N \otimes \mathcal{O}/\varpi^n$ where N is admissible and flat over $\mathcal{O}$ and $N \otimes_{\mathcal{O}} \overline{K}$ is semi-simple. The (implicit) assumptions on $\overline{\rho}$ in [3] are, however, somewhat weaker; one only need assume that the particular eigenvalue α has multiplicity one, and moreover the assumption that M is an element of $\mathscr{C}$ is relaxed (although, for the module M in Sect. 3 of [3] for which Lemma 3.2.2 is applied, one may deduce from local-global compatibility that $M \in \mathscr{C}$). We expect that Theorem 9.16 is true under these weaker assumptions as well, and possibly even under the generalization of Lemma 3.2.2 of [3] due to Thorne (Proposition 5.9 of [53]), see Remark 9.18 following the proof.

Proof For $F = \mathscr{F}$ or $\mathscr{G}$, one has $F(M) = \varinjlim F(M_i)$ as the limit runs over all finite length submodules M_i, hence it suffices to prove the isomorphism for M of finite length. In particular, we may assume that $M = \text{n-ind}_B^G M_B$ for some finite deformation $\widetilde{\chi}$ of χ. Then we have an isomorphism

$$\mathscr{F}(M) \simeq (M_B)^{D(\mathcal{O}_x)}.$$

Let $\widetilde{\chi}(m)$ denote the restriction of $\widetilde{\chi}$ to $F_x^\times$ whose irreducible constituents correspond to the unramified character which takes the value α_m on a uniformizer for some eigenvalue α_m of $\overline{\rho}(\mathrm{Frob}_x)$. Let $\widetilde{\chi}(\widehat{m})$ denote the restriction of $\widetilde{\chi}$ to $(F_x^\times)^{n-1}$ corresponding to the other $n-1$ eigenvalues. Then there is an isomorphism

$$\begin{aligned} M^{U_0(x)} &\simeq \bigoplus_{m=1}^{n} (\widetilde{\chi}(m) \otimes \text{n-ind}(\widetilde{\chi}(\widehat{m})))^{L(\mathcal{O}_x)} \\ &\simeq \bigoplus_{m=1}^{n} (M_B)^{D(\mathcal{O}_x)}. \end{aligned}$$

Moreover, the action of V on each factor is given by the coset corresponding to $\varpi_x \times \mathrm{Id} \in L(F_x)$, which acts via $\widetilde{\chi}(m)(\varpi_x)$, and the normalized sum of the invariants is equal to the image of $M^{G(\mathcal{O}_x)}$. In particular, the operator e_α projects onto the mth factor such that $\alpha = \alpha_m$, which is an isomorphism. □

Remark 9.18 The proof of this result, is, to some extent, "by explicit computation." Here is a different approach which may work under the weaker assumption that α has multiplicity one but there is no other assumption on the eigenvalues of $\overline{\rho}(\mathrm{Frob}_x)$. First one establishes, for irreducible $\pi \in \mathscr{C}$, that there is an isomorphism $\mathscr{F}(\pi) \simeq \mathscr{G}(\pi)$. This is essentially already done in Sect. 3 of [3]. Now proceed by induction on the length of M. Suppose now that the claim is true for modules of length $<$ length(M). By assumption, there is an inclusion $\pi \subset M$, let N denote the quotient. By induction, there is a long exact sequence as follows:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mathscr{F}(\pi) & \longrightarrow & \mathscr{F}(M) & \longrightarrow & \mathscr{F}(N) & \longrightarrow & R^1\mathscr{F}(\pi) \\ & & \| & & \downarrow & & \| & & \downarrow \\ 0 & \longrightarrow & \mathscr{G}(\pi) & \longrightarrow & \mathscr{G}(M) & \longrightarrow & \mathscr{G}(N) & \longrightarrow & R^1\mathscr{G}(\pi) \end{array}$$

By the five lemma, it suffices to show that $R^1\mathscr{F}(\pi) \to R^1\mathscr{G}(\pi)$ is injective. By Hochschild–Serre, one has isomorphisms

$$R^1\mathscr{F}(\pi) \simeq H^1(\mathrm{GL}_n(k), \pi^{U(x)}), \qquad R^1\mathscr{G}(\pi) \simeq e_\alpha H^1(U_0(k), \pi^{U(x)}),$$

which (in principle) one might be able to compute explicitly for the relevant π.

9.4.2 *Applications to Taylor–Wiles primes*

We now define the modules M_j as follows:

$$M_j := \lim_{m\to\infty} H^j(X(x^m), \mathcal{O}/\varpi^n)_{\mathfrak{m}}.$$

The modules M_j are filtered (as $G = \mathrm{GL}_n(F_x)$-modules) by the the admissible module $M_j[\mathfrak{m}]$. By assumption, any representation $\pi \subset M_j[\mathfrak{m}]$ lies in $\mathscr{C}$, and hence $M_j \in \mathscr{C}$ by Lemma 9.10. By Hochschild–Serre, we have two spectral sequences, namely,

$$H^i(G(\mathcal{O}_x), M_j) \Rightarrow H^{i+j}(X, \mathcal{O}/\varpi^n)_{\mathfrak{m}_\emptyset},$$
$$e_\alpha H^i(U_0(x), M_j) \Rightarrow e_\alpha H^{i+j}(X_0(x), \mathcal{O}/\varpi^n)_{\mathfrak{m}_\emptyset} = H^{i+j}(X_0(x), \mathcal{O}/\varpi^n)_{\mathfrak{m}}.$$

There is a natural map between these spectral sequences given by ι_*. By Theorem 9.16, these maps are isomorphisms, and hence we deduce that the map:

$$e_\alpha : H^*(X, \mathcal{O}/\varpi^n)_{\mathfrak{m}_\emptyset} \simeq H^*(X_0(x), \mathcal{O}/\varpi^n)_{\mathfrak{m}},$$

is an isomorphism, as required.

9.5 Modularity lifting

In this section we prove our main theorem on modularity lifting. We note that Theorem 5.16 follows from it immediately as a corollary.

We assume the existence of a maximal ideal $\mathfrak{m}$ of $\mathbf{T} := \mathbf{T}_{\emptyset,1}$ with $\overline{r}_{\mathfrak{m}} \cong \overline{r}$. We assume also that $\overline{r}(G_{F(\zeta_p)})$ is big.

For each integer $N \geq 1$, let Q_N be a set of primes satisfying the conclusions of Proposition 8.5. We also assume that Conjecture B holds for each of the sets Q_N.

For each N, there is a natural covering map $Y_1(Q_N) \to Y_0(Q_N)$ with Galois group

$$\widetilde{\Delta} := \prod_{x \in Q} (\mathcal{O}_F/x)^{\times}.$$

Choose a surjection $\widetilde{\Delta} \twoheadrightarrow \Delta_N := (\mathbf{Z}/p^N\mathbf{Z})^q$ and let $Y_{\Delta_N}(Q_N) \to Y_0(Q_N)$ be the corresponding sub-cover. For each $0 \leq M \leq N$, we regard Δ_M as a quotient of Δ_N in the natural fashion. This gives rise to further sub-covers $Y_{\Delta_M}(Q_N) \to Y_0(Q_N)$.

By Conjecture B and the results of Sect. 7.1, there exists a perfect complex $\widetilde{D}_N$ of free $S_N = \mathcal{O}[\Delta_N]$-modules such that

- $\widetilde{D}_N$ is concentrated in degrees $q_0, \ldots, q_0 + l_0$,
- the complex $\widetilde{D}_N \otimes_{S_N} S_N/\mathfrak{m}_{S_N}$ has trivial differentials,
- for each $i, n \geq 1$ and $0 \leq M \leq N$, we have an isomorphism of S_N-modules

$$H_i(\widetilde{D}_N \otimes_{S_N} S_M/\varpi^n) \cong \left(\lim_{n \to \infty} \prod_{x \in Q} Q_x(V_x)^{n!} \right) H_i(Y_{\Delta_M}(Q_N), \mathcal{O}/\varpi^n)_{\mathfrak{m}^*}.$$

Note that we have

$$H^*(Y_1(Q_N), \mathcal{O}/\varpi^n)^{\vee}_{\mathfrak{m}} \simeq H_*(Y_1(Q_N), \mathcal{O}/\varpi^n)_{\mathfrak{m}^*},$$

where the equivalence comes from the fact that we are assuming the cohomology of the boundary vanishes after localization at $\mathfrak{m}$.

Similarly, working with the local system associated to our choice of characters $\chi = (\chi_v)_{v \in R}$, there exists a perfect complex $\widetilde{D}_N^{\chi}$ of free S_N-modules satisfying the first two properties above as well as:

- for each $i, n \geq 1$ and $0 \leq M \leq N$, we have an isomorphism of S_N-modules

$$\begin{aligned} &H_i(\widetilde{D}_N^{\chi} \otimes_{S_N} S_M/\varpi^n) \\ &\quad \cong \left(\lim_{n \to \infty} \prod_{x \in Q} Q_x(V_x)^{n!} \right) H_{i,\chi}(Y_{\Delta_M}(Q_N), \mathcal{O}/\varpi^n)_{\mathfrak{m}^*}. \end{aligned}$$

Note that we have

$$H^*_\chi(Y_1(Q_N), \mathcal{O}/\varpi^n)^\vee_\mathfrak{m} \simeq H_{*,\chi}(Y_1(Q_N), \mathcal{O}/\varpi^n)_{\mathfrak{m}^*},$$

again by Conjecture B part (4)).

Since

$$H^*(Y, \mathcal{O}/\varpi) \cong H^*_\chi(Y, \mathcal{O}/\varpi),$$

the ideal $\mathfrak{m}$ induces a maximal ideal of $\mathbf{T}_\chi := \mathbf{T}_{\emptyset,\chi}$, which we also denote by $\mathfrak{m}$ in a slight abuse of notation. By Conjecture B, we have surjections $R_\mathcal{S} \twoheadrightarrow \mathbf{T}_\mathfrak{m}$ and $R_{\mathcal{S}_\chi} \twoheadrightarrow \mathbf{T}_{\chi,\mathfrak{m}}$.

Theorem 9.19 *If we regard $H^{q_0}(Y, K/\mathcal{O})^\vee_\mathfrak{m}$ as an $R_\mathcal{S}$-module via the map $R_\mathcal{S} \twoheadrightarrow \mathbf{T}_\mathfrak{m}$, then it is a nearly faithful $R_\mathcal{S}$-module.*

Proof We will apply the results of Sect. 6.1. For each $N \geq 1$, we have chosen a set of Taylor–Wiles primes Q_N satisfying the assumptions of Proposition 8.5 (for some fixed choice of q). Let

$$g = q + |T| - 1 - [F : \mathbf{Q}]\frac{n(n-1)}{2} - l_0$$

be the integer appearing in part (4) of this proposition. We will apply Proposition 6.6 with the following:

- Let $S_\infty = \mathcal{O}[(\mathbf{Z}_p)^q]$ and $S_N = \mathcal{O}[\Delta_N]$ as in the statement of Proposition 6.3.
- Let $j = n^2|T| - 1$ and $\mathcal{O}^\square = \mathcal{O}[z_1, \ldots, z_j]$.
- Let

$$R^1_\infty = R^1_{\mathrm{loc}}[x_1, \ldots, x_g]$$
$$R^2_\infty = R^\chi_{\mathrm{loc}}[x_1, \ldots, x_g].$$

  Note each R^i_∞ is p-torsion free and equidimensional of dimension $1 + q + j - l_0$ by Lemma 8.2. In addition, we have a natural isomorphism $R^1_\infty/\varpi \xrightarrow{\sim} R^2_\infty/\varpi$.
- Let $(R^1, H^1) = (R_\mathcal{S}, H^{q_0}(Y, K/\mathcal{O})^\vee_\mathfrak{m})$ and $(R^2, H^2) = (R_{\mathcal{S}^\chi}, H^{q_0}_\chi(Y, K/\mathcal{O})^\vee_\mathfrak{m})$. Note that we have natural compatible isomorphisms $R^1/\varpi \xrightarrow{\sim} R^2/\varpi$ and $H^1/\varpi \xrightarrow{\sim} H^2/\varpi$.
- Let $T = T^1 = T^2$ be the complex with $T^i = H^{i-l_0}(Y, \mathcal{O}/\varpi)^\vee_\mathfrak{m}$ and with all differentials $d : T^i \to T^{i+1}$ equal to 0.

- For each $N \geq 1$, let $Y_{\Delta_N}(Q_N) \to Y_0(Q_N)$ denote the subcover of $Y_1(Q_N) \to Y_0(Q_N)$ with Galois group $\Delta_N = (\mathbf{Z}/p^N)^q$. We introduced above perfect (homological) complexes of S_N-modules $\widetilde{D}_N$ and $\widetilde{D}_N^\chi$ above; they are concentrated in degrees $q_0, \ldots, q_0 + l_0$. We now regard these as cohomological complexes concentrated in degrees $0, \ldots, l_0$. Then we let D_N^1 (resp. D_N^2) denote the perfect complex of S_N-modules $\widetilde{D}_N/\varpi^N$ (resp. $\widetilde{D}_N^\chi/\varpi^N$). Note that the cohomology of D_N^1 (resp. D_N^2) computes $H^*(Y_{\Delta_N}(Q_N), \mathcal{O}/\varpi^N)_{\mathfrak{m}}^\vee$ (resp. $H_\chi^*(Y_{\Delta_N}(Q_N), \mathcal{O}/\varpi^N)_{\mathfrak{m}}^\vee$), after a shift in degree by q_0. We can and do assume that $D_N^i \otimes S_N/\mathfrak{m}_{S_N} \cong T$ for $i = 1, 2$.
- Choose representatives for the universal deformations of type $\mathcal{S}_{Q_N}$ and $\mathcal{S}_{Q_N}^\chi$ which agree modulo ϖ. This gives rise to isomorphisms

$$R_{\mathcal{S}_{Q_N}}^{\square_T} \xrightarrow{\sim} R_{\mathcal{S}_{Q_N}}[z_1, \ldots, z_j]$$
$$R_{\mathcal{S}_{Q_N}^\chi}^{\square_T} \xrightarrow{\sim} R_{\mathcal{S}_{Q_N}^\chi}[z_1, \ldots, z_j].$$

  In the notation of Proposition 6.3, the rings on the right hand side can be written $R_{\mathcal{S}_{Q_N}}^\square$ and $R_{\mathcal{S}_{Q_N}^\chi}^\square$. By Proposition 8.5, we can and do choose surjections $R_\infty^1 \twoheadrightarrow R_{\mathcal{S}_{Q_N}}^\square$ and $R_\infty^2 \twoheadrightarrow R_{\mathcal{S}_{Q_N}^\chi}^\square$. Composing these with the natural maps $R_{\mathcal{S}_{Q_N}}^\square \twoheadrightarrow R_{\mathcal{S}_{Q_N}} \twoheadrightarrow R_{\mathcal{S}} = R^1$ and $R_{\mathcal{S}_{Q_N}^\chi}^\square \twoheadrightarrow R_{\mathcal{S}_{Q_N}^\chi} \twoheadrightarrow R_{\mathcal{S}^\chi} = R^2$, we obtain surjections $\phi_N^1 : R_\infty^1 \twoheadrightarrow R^1$ and $\phi_N^2 : R_\infty^2 \twoheadrightarrow R^2$.

We have now introduced all the necessary input data to Proposition 6.6. We now check that they satisfy the required conditions.

- For each $M \geq N \geq 0$ with $M \geq 1$ and each $n \geq 1$, we have an action of $R_{\mathcal{S}_{Q_N}}$ (resp. $R_{\mathcal{S}_{Q_N}^\chi}$) on the cohomology $H^*(Y_{\Delta_N}(Q_M), \mathcal{O}/\varpi^n)_{\mathfrak{m}}$ (resp. $H_\chi^*(Y_{\Delta_N}(Q_M), \mathcal{O}/\varpi^n)_{\mathfrak{m}}$) by Conjecture B. Applying the functor, $X \mapsto X^\square$, and using the surjection $R_\infty^1 \twoheadrightarrow R_{\mathcal{S}_{Q_N}}^\square$ (resp. $R_\infty^2 \twoheadrightarrow R_{\mathcal{S}_{Q_N}^\chi}$), we obtain an action of R_∞^1 (resp. R_∞^2) on $H^*(D_M^{1,\square} \otimes_{S_M} S_N/\varpi^n)$ (resp. $H^*(D_M^{2,\square} \otimes_{S_M} S_N/\varpi^n)$).

Thus condition (b) of Proposition 6.3 is satisfied for both sets of patching data. Condition (d) follows from Conjecture B, while condition (c) is clear. Finally, we note that we have isomorphisms

$$H^{l_0}((D_M^1)^\square \otimes_{S_M} S_N/\varpi) = H^{l_0}(Y_{\Delta_N}(Q_M), \mathcal{O}/\varpi)^\square \xrightarrow{\sim} H_\chi^{l_0}(Y_{\Delta_N}(Q_M), \mathcal{O}/\varpi)^\square$$
$$= H^{l_0}((D_M^2)^\square \otimes_{S_M} S_N/\varpi)$$

for all $M \geq N \geq 0$ with $M \geq 1$. These isomorphisms are compatible with the actions of R_∞^i and give rise to the commutative square required by Proposition 6.6.

We have now satisfied all the requirements of Proposition 6.6 and hence we obtain two complexes $P_\infty^{1,\square}$ and $P_\infty^{2,\square}$. By Lemma 8.2, Spec R_∞^2 is irreducible, and hence by Theorem 6.4 $H^{l_0}(P_\infty^{2,\square})$ is nearly faithful as an R_∞^2-module. Thus

$$H^{l_0}(P_\infty^{1,\square})/\varpi \cong H^{l_0}(P_\infty^{2,\square})/\varpi$$

is nearly faithful over $R_\infty^1/\varpi \xrightarrow{\sim} R_\infty^2/\varpi$. By Lemma 8.2 and [11, Lemma 2.2], it follows that $H^{l_0}(P_\infty^{1,\square})$ is nearly faithful over R_∞^1 providing that $H^{l_0}(P_\infty^{1,\square})$ is p-torsion free. However, each associated prime of $H^{l_0}(P_\infty^{1,\square})$ is a minimal prime of R_∞^1 and by Lemma 8.2, all such primes have characteristic 0. Thus p cannot be a zero divisor on $H^{l_0}(P_\infty^{1,\square})$ and the result of Taylor applies. By conclusion (iv) of Proposition 6.3 we deduce that $H^1 = H^{q_0}(Y, K/\mathcal{O})_{\mathfrak{m}}^\vee$ is nearly faithful over $R^1 = R_{\mathcal{S}}$, as required. □

10 Proof of Theorem 1.1

In this section, we prove Theorem 1.1

Proof Let A be an elliptic curve over a number field K. If A has CM, then the result is well known, so we may assume that $\mathrm{End}_{\mathbf{C}}(A) = \mathbf{Z}$. Let

$$r = \mathrm{Sym}^{2n-1}\rho : G_K \to \mathrm{GL}_{2n}(\mathbf{Q}_p)$$

denote the representation corresponding to the $(2n-1)$th symmetric power of the Tate module of A. To prove Theorem 1.1, it suffices (following, for example, the proof of Theorem 4.2 of [81]) to prove that for each n, there exists a p such that r is potentially modular. We follow the proof of Theorem 6.4 of [76]. (The reason for following the proof of Theorem 6.4 instead of Theorem 6.3 of *ibid*. is that the latter theorem proceeds via compatible families arising from the Dwork family such that $V[\lambda]_t$ is ordinary but not crystalline, which would necessitate a different version of Theorem 5.16). In particular, we make the following extra hypothesis:

- There exists a prime p which is totally split in K, and such that $p+1$ is divisible by an integer N_2 which is greater than n and prime to the conductor of A. Moreover, the mod-p representation $\overline{\rho}_A : G_K \to \mathrm{GL}_2(\mathbf{F}_p)$ associated to $A[p]$ is surjective, A has good reduction at all $v|p$, and for all primes $v|p$ we have

$$\overline{\rho}_A : G_{\mathbf{Q}_p} \simeq \mathrm{Ind}_{\mathbf{Q}_{p^2}}^{\mathbf{Q}_p} \omega_2.$$

(This is a non-trivial condition on A, we consider the general case below). It then suffices to find sufficiently large primes p and l, a finite extension L/K, an integer N_2 with $N_2 > n + 1$ and $p + 1 = N_1 N_2$ (as in the statement of Theorem 6.4 of Sect. 4 of [76]) and primes λ, λ' of $\mathbf{Q}(\zeta_N)^+$ (with λ dividing p and λ' dividing l) and a point $t \in T_0(L)$ on the Dwork family such that:

(1) $V[\lambda]_t \simeq \overline{r}|_{G_L}$,
(2) $V[\lambda']_t \simeq \overline{r}'|_{G_L}$, where r' is an ordinary weight 0 representation which induced from G_{LM} for some suitable CM field $M/\mathbf{Q}$ of degree $2n$.
(3) p splits completely in L.
(4) A and V are semistable over L.
(5) $\overline{r}|_{G_L}$ and $\overline{r}'|_{G_L}$ satisfy all the hypotheses of Theorem 5.16 with the possible exception of residual modularity.

This can be deduced (as in the proof of Theorem 6.4 of [76]) using the theorem of Moret–Bailly (in the form of Proposition 6.2 of *ibid*) and via character building. By construction, the modularity of r follows from two applications of Theorem 5.16, once applied to the λ'-adic representation associated to V (using $\overline{r}'$ and the residual modularity coming from the induction of a Grossencharacter) and once to the λ-adic representation associated to $\mathrm{Sym}^{2n-1}(A)$, using the residual modularity coming from V.

For a general elliptic curve E, we reduce to the previous case as follows. It suffices to find a second elliptic curve A, a number field L/K, and primes p and q such that:

(1) The mod-p representation $\overline{r} = (\mathrm{Sym}^{2n-1}\overline{\rho}_E)|_{G_L}$ satisfies all the hypotheses of Theorem 5.16 with the possible exception of residual modularity.
(2) A and E are semistable over L and have good reduction at all primes dividing p and q.
(3) p and q split completely in L.
(4) $p+1$ is divisible by an integer $N_2 > n+1$ which is prime to the conductor of A.
(5) $E[q] \simeq A[q]$ as G_L-modules, and the corresponding mod-p representation is surjective.
(6) The mod-p representation $\overline{\rho}_A : G_L \to \mathrm{GL}_2(\mathbf{F}_p)$ associated to A is surjective, and $\overline{\rho}_A|_{G_{\mathbf{Q}_p}} \simeq \mathrm{Ind}^{\mathbf{Q}_p}_{\mathbf{Q}_{p^2}} \omega_2$.

This lemma also follows easily from Proposition 6.2 of [76], now applied to twists of a modular curve. We deduce as above (using the mod-q representation) that $\mathrm{Sym}^{2n-1}(A)$ is potentially modular over some extension which is unramified at p, and then use Theorem 5.16 once more now at the prime at p to deduce that $\mathrm{Sym}^{2n-1}(E)$ is modular. □

Remark 10.1 It is no doubt possible to also deal with even symmetric powers using the tensor product idea of Harris [82].

Acknowledgements We would like to thank Matthew Emerton, David Helm, Toby Gee, Andrew Snowden, Richard Taylor, David Treumann, Akshay Venkatesh, and Andrew Wiles for useful conversations. We also thank Toby Gee for detailed comments on several drafts of this paper. Many of the ideas of this paper were discovered when both the authors were members of the Institute for Advanced Study in 2010–2011—the authors would like to thank the IAS for the pleasant and productive environment provided. Finally, the authors would like to thank each of the referees for their help in improving both the accuracy and readability of this manuscript.

References

1. Wiles, A.: Modular elliptic curves and Fermat's last theorem. Ann. Math. (2) **141**(3), 443–551 (1995)
2. Taylor, R., Wiles, A.: Ring-theoretic properties of certain Hecke algebras. Ann. Math. (2) **141**(3), 553–572 (1995)
3. Clozel, L., Harris, M., Taylor, R.: Automorphy for some l-adic lifts of automorphic mod l Galois representations. Publ. Math. Inst. Hautes Études Sci. (108), 1–181 (2008) With Appendix A, summarizing unpublished work of Russ Mann, and Appendix B by Marie-France Vignéras
4. Calegari, F., Emerton, M.: Completed cohomology—a survey. Non-abelian fundamental groups and Iwasawa theory, London Math. Soc. Lecture Note Ser., vol. 393, pp. 239–257, Cambridge University Press, Cambridge (2012)
5. Calegari, F., Mazur, B.: Nearly ordinary Galois deformations over arbitrary number fields. J. Inst. Math. Jussieu **8**(1), 99–177 (2009)
6. Calegari, F., Venkatesh, A.: Towards a Torsion Jacquet–Langlands correspondence. Preprint
7. Harris, M., Lan, K.-W., Taylor, R., Thorne, J.: On the rigid cohomology of certain Shimura varieties. Res. Math. Sci. 3 (2016), Paper No. 37, 308
8. Scholze, P: On torsion in the cohomology of locally symmetric varieties. Ann. Math. (2) **182**(3), 945–1066 (2015)
9. Borel, A., Wallach, N.: Continuous cohomology, discrete subgroups, and representations of reductive groups, second ed., Mathematical Surveys and Monographs, vol. 67, American Mathematical Society, Providence (2000)
10. Harris, M.: Automorphic forms and the cohomology of vector bundles on Shimura varieties, Automorphic forms, Shimura varieties, and L-functions, Vol. II (Ann Arbor, MI, 1988). Perspect. Math., vol. 11., pp. 41–91. Academic, Boston (1990)
11. Taylor, R.: Automorphy for some l-adic lifts of automorphic mod l Galois representations. II. Publ. Math. Inst. Hautes Études Sci. **108**, 183–239 (2008)
12. Fontaine, J.-M.: Représentations p-adiques semi-stables, Astérisque (1994), vol. 223, pp. 113–184, With an appendix by Pierre Colmez, Périodes p-adiques (Bures-sur-Yvette, 1988)
13. Diamond, F.: On deformation rings and Hecke rings. Ann. Math. (2) **144**(1), 137–166 (1996)
14. Diamond, F.: An extension of Wiles' results, Modular forms and Fermat's last theorem (Boston, MA, 1995), pp. 475–489. Springer. New York, 1997 (1995)
15. Conrad, B., Diamond, F., Taylor, R.: Modularity of certain potentially Barsotti–Tate Galois representations. J. Am. Math. Soc. **12**(2), 521–567 (1999)
16. Pilloni, V.: Modularité, formes de Siegel et surfaces abéliennes. J. Reine Angew. Math. **666**, 35–82 (2012)
17. Harris, M.: The Taylor–Wiles method for coherent cohomology. J. Reine Angew. Math. **679**, 125–153 (2013)
18. Lan, K.-W., Suh, J.: Vanishing theorems for torsion automorphic sheaves on general PEL-type Shimura varieties. Adv. Math. **242**, 228–286 (2013)

19. Deligne, P., Serre, J.-P.: Formes modulaires de poids 1. Ann. Sci. École Norm. Sup. (4) **7**(1974), 507–530 (1975)
20. Goldring, W., Shin, S.W.: Galois representations associated to holomorphic limits of discrete series. Compos. Math. **150**(2), 191–228 (2014)
21. Edixhoven, B.: Comparison of integral structures on spaces of modular forms of weight two, and computation of spaces of forms mod 2 of weight one. J. Inst. Math. Jussieu **5**(1), 1–34 (2006). With appendix A (in French) by Jean-François Mestre and appendix B by Gabor Wiese
22. Buzzard, K., Taylor, R.: Companion forms and weight one forms. Ann. Math. (2) **149**(3), 905–919 (1999)
23. Buzzard, K.: Analytic continuation of overconvergent eigenforms. J. Am. Math. Soc. **16**(1), 29–55 (2003). (electronic)
24. Khare, C., Wintenberger, J.-P.: Serre's modularity conjecture. I. Invent. Math. **178**(3), 485–504 (2009)
25. Boston, N.: Some cases of the Fontaine–Mazur conjecture. II. J. Number Theory **75**(2), 161–169 (1999)
26. Diamond, F.: The Taylor–Wiles construction and multiplicity one. Invent. Math. **128**(2), 379–391 (1997)
27. Henri, D., Fred, D., Richard, T.: Fermat's last theorem, Elliptic curves, modular forms & Fermat's last theorem (Hong Kong, 1993), pp. 2–140. Int. Press, Cambridge (1997)
28. Gross, B.H.: A tameness criterion for Galois representations associated to modular forms (mod p). Duke Math. J. **61**(2), 445–517 (1990)
29. Katz, N.M.: A result on modular forms in characteristic p, Modular functions of one variable, V (Proc. Second Internat. Conf., Univ. Bonn, Bonn, 1976). Lecture Notes in Math., vol. 601, pp. 53–61. Springer, Berlin (1977)
30. Conrad, B.: Arithmetic moduli of generalized elliptic curves. J. Inst. Math. Jussieu **6**(2), 209–278 (2007)
31. Edixhoven, B.: The weight in Serre's conjectures on modular forms. Invent. Math. **109**(3), 563–594 (1992)
32. Hartshorne, R.: Residues and duality, Lecture notes of a seminar on the work of A. Grothendieck, given at Harvard 1963/64. With an appendix by P. Deligne. Lecture Notes in Mathematics, No. 20, Springer, Berlin (1966)
33. Milne, J.S.: Étale Cohomology, Princeton Mathematical Series, vol. 33. Princeton University Press, Princeton (1980)
34. Ribet, K.A.: Irreducible Galois representations arising from component groups of Jacobians, Elliptic curves, modular forms, & Fermat's last theorem (Hong Kong, 1993). Ser. Number Theory, I, pp. 131–147. Int. Press, Cambridge (1995)
35. Serre, J.-P.: Letter to K. Ribet (1987) (unpublished)
36. Serre, J.-P.: Sur les représentations modulaires de degré 2 de Gal$(\overline{Q}/Q)$. Duke Math. J. **54**(1), 179–230 (1987)
37. Robert, F.: Coleman and José Felipe Voloch, Companion forms and Kodaira–Spencer theory. Invent. Math. **110**(2), 263–281 (1992)
38. Wiese, G.: On Galois representations of weight one. Doc. Math. **19**, 689–707 (2014)
39. Carayol, H.: Formes modulaires et représentations galoisiennes à valeurs dans un anneau local complet, p-adic monodromy and the Birch and Swinnerton–Dyer conjecture (Boston, MA, 1991), Contemp. Math., vol. 165, pp. 213–237, Amer. Math. Soc., Providence, 1994 (1991)
40. Geraghty, D.: Modularity lifting theorems for ordinary galois representations. Preprint
41. Snowden, A.: Singularities of ordinary deformation rings. Math. Zeit. (to appear)
42. Deligne, P., Rapoport, M.: Les schémas de modules de courbes elliptiques, Modular functions of one variable, II (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972), pp. 143–316. Lecture Notes in Math., Vol. 349. Springer, Berlin (1973)

43. Mumford, D.: Abelian varieties, Tata Institute of Fundamental Research Studies in Mathematics, vol. 5, Published for the Tata Institute of Fundamental Research, Bombay, 2008, With appendices by C. P. Ramanujam and Yuri Manin, Corrected reprint of the second edition (1974)
44. Katz, N.M., Mazur, B.: Arithmetic Moduli of Elliptic Curves, Annals of Mathematics Studies, vol. 108. Princeton University Press, Princeton (1985)
45. Eisenbud, D: Commutative algebra, Graduate Texts in Mathematics, vol. 150, Springer, New York, 1995, With a view toward algebraic geometry
46. Herzog, J., Kunz, E. (eds).: Der kanonische Modul eines Cohen-Macaulay-Rings, Lecture Notes in Mathematics, Vol. 238, Springer, Berlin, 1971, Seminar über die lokale Kohomologietheorie von Grothendieck, Universität Regensburg, Wintersemester 1970/1971
47. Rossi, M.: Hilbert Functions of Cohen–Macaulay Local Rings, Notes of the Lectures at the School in Commutative Algebra and Its Connections to Geometry, Olinda (2009)
48. Mazur, B.: Modular curves and the Eisenstein ideal. Inst. Hautes Études Sci. Publ. Math. **47**, 33–186 (1977/1978)
49. Mazur, B., Ribet, K.A.: Two-dimensional representations in the arithmetic of modular curves. Astérisque **196–197**(6), 215–255 (1991/1992). Courbes modulaires et courbes de Shimura (Orsay, 1987/1988)
50. Ribet, K.A., Stein, W.A.: Lectures on Serre's conjectures, Arithmetic algebraic geometry (Park City, UT, 1999), IAS/Park City Math. Ser., vol. 9, pp. 143–232. Amer. Math. Soc., Providence (2001). With an appendix by Kevin Buzzard
51. Wiese, G.: Multiplicities of Galois representations of weight one. Algebra Number Theory **1**(1), 67–85 (2007). With an appendix by Niko Naumann
52. Kisin, M.: Moduli of finite flat group schemes, and modularity. Ann. Math. (2) **170**(3), 1085–1180 (2009)
53. Thorne, J.: On the automorphy of l-adic Galois representations with small residual image. J. Inst. Math. Jussieu (2011) (to appear)
54. Kisin, M.: Modularity of 2-dimensional Galois representations, Current developments in mathematics, 2005, pp. 191–230. Int. Press, Somerville (2007)
55. Reduzzi, D.A.: On the number of irreducible components of universal deformation rings in the $l \neq p$ case (2013). http://msp.org/extras/D.A.Reduzzi-Irreducible_Components_Local_Deformations.pdf
56. Carayol, H.: Sur les représentations galoisiennes modulo l attachées aux formes modulaires. Duke Math. J. **59**(3), 785–801 (1989)
57. Edixhoven, B.: Serre's conjecture, Modular forms and Fermat's last theorem (Boston, MA, 1995), pp. 209–242. Springer. New York (1997)
58. Mazur, B.: Deforming Galois representations, Galois groups over Q (Berkeley, CA, 1987), Math. Sci. Res. Inst. Publ., vol. 16, pp. 385–437. Springer, New York (1989)
59. Grunewald, F., Helling, H., Mennicke, J.: SL_2 over complex quadratic number fields. I. Algebra Log. **17**(5), 512–580 (1978). 622
60. Fritz, G.: Notes on Bianchi manifolds (1972) (unpublished manuscript)
61. Ash, A.: Galois representations attached to mod p cohomology of $\mathrm{GL}(n, Z)$. Duke Math. J. **65**(2), 235–255 (1992)
62. Figueiredo, L.M.: Serre's conjecture for imaginary quadratic fields. Compos. Math. **118**(1), 103–122 (1999)
63. Taylor, R.L.: On congruences between modular forms. Ph.D. Thesis, ProQuest LLC, Princeton University, Ann Arbor (1988)
64. Harder, G.: Eisenstein cohomology of arithmetic groups. The case GL_2. Invent. Math. **89**(1), 37–118 (1987)
65. Berger, T.T.: An Eisenstein ideal for imaginary quadratic fields. Ph.D. Thesis, ProQuest LLC, Ann Arbor, MI, University of Michigan (2005)

66. Calegari, F.: Even Galois representations and the Fontaine–Mazur conjecture. Invent. Math. **185**(1), 1–16 (2011)
67. Skinner, C.M., Wiles, A.J.: Base change and a problem of Serre. Duke Math. J. **107**(1), 15–25 (2001)
68. Tilouine, J.: Nearly ordinary rank four Galois representations and p-adic Siegel modular forms. Compos. Math. **142**(5), 1122–1156 (2006). With an appendix by Don Blasius
69. Matsumura, H.: Commutative ring theory, Cambridge Studies in Advanced Mathematics, vol. 8, Cambridge University Press, Cambridge (1986). Translated from the Japanese by M. Reid
70. The Stacks Project Authors, stacks project (2014). http://stacks.math.columbia.edu
71. Lan, K.-W.: Arithmetic compactifications of PEL-type Shimura varieties. Ph.D. Thesis, ProQuest LLC, Harvard University, Ann Arbor (2008)
72. Nakajima, S.: On Galois module structure of the cohomology groups of an algebraic variety. Invent. Math. **75**(1), 1–8 (1984)
73. Kaplansky, I.: Projective modules. Ann. Math. (2) **68**, 372–377 (1958)
74. Lan, K.-W.: Toroidal compactifications of PEL-type Kuga families. Algebra Number Theory **6**(5), 885–966 (2012)
75. Faltings, G.Chai, C.-L.: Degeneration of abelian varieties, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 22, Springer, Berlin (1990), With an appendix by David Mumford
76. Barnet-Lamb, T., Geraghty, D., Harris, M., Taylor, R.: A family of Calabi–Yau varieties and potential automorphy II. Publ. Res. Inst. Math. Sci. **47**(1), 29–98 (2011)
77. Bass, H., Milnor, J., Serre, J.-P.: Solution of the congruence subgroup problem for SL_n ($n \geq 3$) and Sp_{2n} ($n \geq 2$). Inst. Hautes Études Sci. Publ. Math. **33**, 59–137 (1967)
78. Serre, J.-P., Arbres, amalgames, SL_2, Société Mathématique de France, Paris, 1977, Avec un sommaire anglais, Rédigé avec la collaboration de Hyman Bass, Astérisque, No. 46
79. Vignéras, M.-F.: Correspondance de Langlands semi-simple pour $\mathrm{GL}(n, F)$ modulo $\ell \neq p$. Invent. Math. **144**(1), 177–223 (2001)
80. Emerton, M., Helm, D.: The local Langlands correspondence for GL_n in families. Ann. Sci. Éc. Norm. Supér. (4) **47**(4), 655–722 (2014)
81. Harris, M., Shepherd-Barron, N., Taylor, R.: A family of Calabi–Yau varieties and potential automorphy. Ann. Math. (2) **171**, 779–813 (2010)
82. Harris, M.: Potential automorphy of odd-dimensional symmetric powers of elliptic curves and applications, Algebra, arithmetic, and geometry: in honor of Yu. I. Manin. Vol. II, Progr. Math., vol. 270, , pp. 1–21. Birkhäuser, Boston (2009)